

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 00:55:42 ; Search time 720.588 Seconds  
(without alignments)  
8741.717 Million cell updates/sec

Title: US-10-814-858A-2

Perfect score: 130

Sequence: 1 tcactatataggaagttc.....ttacatctagataaacaatg 130

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130	100.0	452	6	B64760
2	130	100.0	452	6	AR427842
3	89	68.5	89	6	B64772
4	89	68.5	89	6	AR427850
5	63.6	48.9	9937	12	AY560325
6	63.6	48.9	9946	12	AY560326
7	62	47.7	735	12	AY183361
8	62	47.7	827	12	AY192160
9	62	47.7	15382	12	AY178047
10	62	47.7	16368	12	AY178049
11	62	47.7	16520	12	AY178048
12	61	46.9	1006	6	A94910
13	61	46.9	1006	6	BD134819
14	60.4	46.5	1138	6	A59870
15	60.4	46.5	1138	6	A94795
16	60	46.2	12607	12	AY737283
17	58.2	44.8	2200	6	AX451740
18	56.2	43.2	5033	6	I28266
19	52	40.0	275	14	TOTMV6

20	52	40.0	6395	6	AR173320	AR173320 Sequence
21	52	40.0	6395	6	AR271575	AR271575 Sequence
22	52	40.0	6395	6	AX040174	AX040174 Sequence
23	52	40.0	6395	6	AX098414	AX098414 Sequence
24	52	40.0	6395	6	AX194392	AX194392 Sequence
25	52	40.0	6395	14	AF165190	AF165190 Tobacco m
26	52	40.0	6395	14	AF273221	AF273221 Tobacco m
27	52	40.0	6395	14	AF395127	AF395127 Tobacco m
28	52	40.0	6395	14	AF395128	AF395128 Tobacco m
29	52	40.0	6395	14	AF395129	AF395129 Tobacco m
30	52	40.0	6395	14	TWV11933	AJ011933 Tobacco m
31	52	40.0	6395	14	TWVCG	X68110 Tobacco mos
32	52	40.0	6395	14	TOTMV4	V01408 Tobacco mos
33	52	40.0	6425	6	AR173322	AR173322 Sequence
34	52	40.0	6425	6	AX098416	AX098416 Sequence
35	52	40.0	6439	6	AR173321	AR173321 Sequence
36	52	40.0	6439	6	AX098415	AX098415 Sequence
37	52	40.0	6446	6	AR173324	AR173324 Sequence
38	52	40.0	6446	6	AX098418	AX098418 Sequence
39	52	40.0	6475	6	AR173323	AR173323 Sequence
40	52	40.0	6475	6	AX098417	AX098417 Sequence
41	52	40.0	7926	6	BD235479	BD235479 Method fo
42	52	40.0	7926	6	AR224869	AR224869 Sequence
43	52	40.0	7926	6	AR397616	AR397616 Sequence
44	52	40.0	11641	6	AX466963	AX466963 Sequence
45	51.8	39.8	1825	6	AR042908	AR042908 Sequence

#### ALIGNMENTS

RESULT 1

E64760

LOCUS  
DEFINITION

E64760 452 bp DNA linear PAT 31-JAN-2002  
Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same.

ACCESSION

E64760 GI:18628523

VERSION

JP 2000139477-A/1.

KEYWORDS

synthetic construct

SOURCE

synthetic construct

ORGANISM

other sequences; artificial sequences.

REFERENCE

1 (bases 1 to 452)

AUTHORS

Tori, R., Sawanto, S.B., Synge, P.K. and Gupta, S.K.

TITLE

Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same  
Patent: JP 2000139477-A 1 23-MAY-2000;  
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH

JOURNAL

OS Artificial Sequence

COMMENT

PN JP 2000139477-A/1

PD 23-MAY-2000

PF 27-APR-1999 JP 1999119227

PR 09-NOV-1998 IN 3322/98

PI RAKESH TORI, SALLY BISHWANATO SAWANTO, PURAJUNNA KUMAR SYNGE,

 PC | C12N15/09, C12N5/10, C12N15/00, C12N5/00 || CC | SHIFU KUMAR GUPUTA |
FT	Key Location/Qualifiers
FT	source 1.452
FT	Location/Qualifiers
FT	1.452
FT	/organism='Artificial Sequence'
FT	/organism='synthetic construct'
FT	/mol\_type='genomic DNA'
FT	/db\_xref='taxon:32630'

ORIGIN

Query Match 100.0%; Score 130; DB 6; Length 452;

Best Local Similarity 100.0%; Pred. No. 6.9e-20;

Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACATATATAGGAAGTTCAATTCATTTTGAATGACACGTTGTTGTCATTTCTCAACAA 60



```

TITLE      Binary expression vector for use in plants
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 9937)
AUTHORS    Ouwerkerk,P.B.F. and Meijer,A.H.
TITLE      Direct Submission
JOURNAL    Submitted (26-FEB-2004) Molecular Cell Biology, Institute of
           Biology, Leiden University, Wassenaarseweg 64, Leiden 2333 AL, The
           Netherlands
FEATURES   Location/Qualifiers
            source
              1..9937
                /organism="Cloning vector pC1300intB-35SnosEX"
                /mol_type="other DNA"
                /db_xref="taxon:267430"
                /notes="derived from pC1300intB (Genbank Accession Number
                AF294977)"
            terminator
              complement(52..303)
                /note="3' NOS; terminator signal of nopaline synthase gene"
            misc_feature
              complement(322..347)
                /note="multiple cloning site; SpeI, XhoI, EcoRI, Sall are
                unique"
            promoter
              complement(425..789)
                /note="CamV 35S promoter"
            misc_feature
              1066..1091
                /note="right border T-DNA repeat"
            misc_feature
              complement(2132..3132)
                /note="STa region from pVS1 plasmid"
            rep_origin
              complement(3725..4725)
                /note="pVS1-REP; replication origin from pVS1"
            misc_feature
              complement(5135..5395)
                /note="bom site from pBR322"
            rep_origin
              complement(5535..5815)
                /note="pBR322 origin of replication"
            gene
              complement(6106..6900)
                /gene="aadA"
            CDS
              complement(6106..6900)
                /gene="aadA"; kanamycin resistance gene from pIG121Hm"
                /codon_start=1
                /transl_table=11
                /product="aminoglycoside transferase"
                /protein_id="AA573283.1"
                /db_xref="GI:45645332"
                /translation="MAKMRISPELKKLIEKYRCVCKTEGMSPAKVYKLVGENENLYLK
                MTD5RYGTTYDVERKDWLWLEKLPVPLVLFERHDSWNLMSADGVLCSEY
                EDQSPKIELVAECIRLPHSIDSCPTVNSLDRSLAELVDVLLNLDVDCENWE
                EDTPKDPRLYDLFKTEPEELVESHGDLGDSNIFVKDGVSGFIDLGRGRADKW
                YDIACVRSIREDIGEEQVELFDFLLGKPKDWEKIKYLLDELDF"
                7325..7350
                /notes="left border repeat from C58 T-DNA"
                /complement(7417..7936)
            promoter
              /note="CamV 35S promoter"
            3'UTR
              complement(7417..7636)
            gene
              /note="CamV polyA signal"
              complement(7652..8867)
            CDS
              complement(join(7652..8338,8529..8867))
                /gene="hptII"
                /note="hptII"
                /codon_start=1
                /transl_table=11
                /product="hygromycin phosphotransferase"
                /protein_id="AA573284.1"
                /db_xref="GI:45645333"
                /translation="MKKPELTATSVKFLIEKFDSDVSLMQLSEGESRAFSDFVGGRR
                GYIVRVNSCAGFYKORYVIRHPASALPIPEVLDDIGFESLTYICISRAAGVTLQD
                LPTELPALQVPAEAMDAIADLSQTPGPGFGQIGQYTTWRDFTCAIDPHVY
                HMQTVMDTVASVAQALDELMAEDCEPVRHLVHADFGSNVLTNDNGRIATAVDWS
                EAMFGDSQVEVANIFFWRPWLACMEQOTRYFERRHPELAGSPLRAYMLRIGLDOLYO
                SLVDGNFDDAANAQGRCDALVRSGAGTVGRTQIARRSAAWTDGCVFVLADSGNRPS
                TRPRACK"
                complement(7652..8338)
            exon
              /gene="hptII"
              /number=2
            intron
              complement(8339..8534)
                /gene="hptII"
                /notes="derived from Ricinus communis catalase-1 intron
                presented in Genbank Accession Number D21161; confers
                higher resistance towards hygromycin in transgenic rice
                and allows easier killing of Agrobacterium tumefaciens
                during regeneration of transgenic plants"
                complement(8529..8867)
                /gene="hptII"
                /number=1
            ORIGIN
              Query Match      48.9%; Score 63.6; DB 12; Length 9937;
              Best Local Similarity 87.7%; Pred. No. 5.7e-05;
              Matches 93; Conservative 0; Mismatches 9; Indels 4; Gaps 2;
              QY      12 AGGAAGTTTCATTTCATTGGAGATGGACA--CGTGTTCATTTCTCAACAATTACCAACA 69
                  |||||
              Db      451 AGGAAGTTTCATTTCATTGGAGAGGACAGGCTTCTTGAGATCCTTCAACAATTACCAACA 392
                  |||||
              QY      70 ACAACAACAACAACAACAACATTATACAAATTACTATTATTAACAATTACA 115
                  |||||
              Db      391 ACAACAACAACAACAACAACAT--TACAATTACTATTATTAACAATTACA 348
                  |||||
            RESULT 6
            LOCUS      AY560326 9946 bp DNA circular SYN 28-MAR-2004
            DEFINITION Cloning vector pC1300intB-35SnosBK, complete sequence.
            ACCESSION  AY560326
            VERSION     AY560326.1 GI:45645334
            KEYWORDS    Cloning vector pC1300intB-35SnosBK
            SOURCE      Cloning vector pC1300intB-35SnosBK
            ORGANISM    Cloning vector pC1300intB-35SnosBK
            REFERENCE   1 (bases 1 to 9946)
            AUTHORS     Ouwerkerk,P.B.F. and Meijer,A.H.
            TITLE       Binary expression vector for use in plants
            JOURNAL      Unpublished
            REFERENCE   2 (bases 1 to 9946)
            AUTHORS     Ouwerkerk,P.B.F. and Meijer,A.H.
            TITLE       Direct Submission
            JOURNAL      Submitted (26-FEB-2004) Molecular Cell Biology, Institute of
            Biology, Leiden University, Wassenaarseweg 64, Leiden 2333 AL, The
            Netherlands
            FEATURES     Location/Qualifiers
                      1..9946
                        /organism="Cloning vector pC1300intB-35SnosBK"
                        /mol_type="other DNA"
                        /db_xref="taxon:267431"
                        complement(48..239)
            terminator
              /note="3' NOS; terminator signal of nopaline synthase gene"
            misc_feature
              complement(318..356)
                /note="multiple cloning site; Sall, EcoRI, XhoI, BglII,
                KpnI, SpeI are unique"
            promoter
              complement(434..798)
                /note="CamV 35S promoter"
            misc_feature
              1075..1100
                /note="right border T-DNA repeat"
            misc_feature
              complement(2141..3141)
                /note="STa region from pVS1 plasmid"
            rep_origin
              complement(3734..4734)
                /note="pVS1-REP; replication origin from pVS1"
            misc_feature
              complement(5144..5404)
                /note="bom site from pBR322"
            rep_origin
              complement(5544..5824)
                /note="pBR322 origin of replication"
            gene
              complement(6115..6909)
                /gene="aadA"
            CDS
              complement(6115..6909)
                /gene="aadA"; kanamycin resistance gene from pIG121Hm"
                /codon_start=1

```

```
/transl_table=11
/product="aminoglycoside transferase"
/protein_id="AAS73285.1"
/db_xref="GI:45645335"
/translation="MAKMRISPKKLIKRYCVKDTGMSPAKVYKLVGENENLYLK
MTDSRYKGYTDVVERKDMWLEGLPVPKVLVFRHDSNLLMSEADGVLCSEY
EDQSPKELIYAEICIRLPHSDIDSCPYTNSLDSRLAELDYLLNNDLADVDGNWE
EDTPFDKPRELDFLTERPEBELVFSHGLDGSNIFVKDGKVSFIDIGRSGRADKW
YDIAFCVRSIREDEEQVELFDLIGIKPWEKIKYLLIDELF"
/misc_feature
complement(7334..7359)
/notes="left border repeat from C58 T-DNA"
/polyA_signal
complement(7426..7645)
/notes="CaMV polyA signal"
gene
complement(7661..8876)
/genes="hptII"
CDS
complement(join(7661..8347,8538..8876))
/genes="hptII"
/notes="hptII"
/codon_start=1
/transl_table=11
/product="hygromycin phosphotransferase"
/protein_id="AAS73286.1"
/db_xref="GI:45645336"
/translation="MKKPELTATSVKFLIEKFDVSVDLMQLSEGESRAFSFDVGG
GVYLRVNSCADGPKDYRVYRHPASAAPIPEVLIDIGFSESLTYCISRAQGVTLQD
LPETELPVLQVPAEMDAIAADLSQTSQGFPGQIGQYTHRDFICADAPHYV
HWQVMDTTSVASVAQALDELMLWAEDECPVRLHVAHDFGSNNVLTDNGRITAVIDWS
EAMFGDSQYEVANIFFWRPWLACMEQOTRYFERRHPELAGSPRLRAYMLRIGLDOLYQ
SLVDGNFDDAAWAQGRCDATVRSAGTGTGRTQARRSAAVWTDGCVVLADSGNRRPS
TRPRANK"
/exon
complement(7661..8347)
/genes="hptII"
/intron
complement(8348..8537)
/genes="hptII"
/notes="derived from Ricinus communis catalase-1 intron
presented in Genbank Accession Number D21161; confers
higher resistance towards hygromycin in transgenic rice
and allows easier killing of Agrobacterium tumefaciens
during regeneration of transgenic plants"
/exon
complement(8538..8876)
/genes="hptII"
/promoter
complement(8923..9703)
/notes="35S promoter from CaMV"
```

```

Query Match      48.9%; Score 63.6; DB 12; Length 9946;
Best Local Similarity 87.7%; Pred. No. 5.7e-05;
Matches 93; Conservative 0; Mismatches 9; Indels 4; Gaps 2;

Qy 12 AGGAAGTTCATTTTCATTTGGAATGGACA--CGTGTGTTCATTTCTCAACAATTACCAACA 69
Db 460 AGGAAGTTCATTTTCATTTGGAATGGACAGGCTTCTTGAGATCTCTCAACAATTACCAACA 401
Qy 70 ACAACAACAACAACAACAATTATACAAATTACTATTATTTACAAATTACA 115
Db 400 ACAACAACAACAACAACAACAT--TACAATTACTATTATTTACAAATTACA 357

RESULT 7
AY183361
LOCUS      AY183361
DEFINITION Synthetic construct duplicated CamV 35/TMV omega-prime leader
sequence fusion promoter, complete sequence.
ACCESSION  AY183361
VERSION     AY183361.1
KEYWORDS    synthetic construct
SOURCE      synthetic construct
ORGANISM    other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 735)
AUTHORS     Yao,Q., Peng,R. and Xiong,A.

Query Match      47.7%; Score 62; DB 12; Length 827;
Best Local Similarity 88.5%; Pred. No. 0.00019;
Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

Qy 12 AGGAAGTTCATTTTCATTTGGAATGGACACGCTGTGTCATTTCTCAACAATTACCAACAAC 71
Db 639 AGGAAGTTCATTTTCATTTGGAATGGACACGCTAT-----TTTACAAACAATTACCAACAAC 693
Qy 72 AACAAACAACAACAACAATTATACAAATTACTATTATTTACAAATTACA 115
Db 694 AACAAACAACAACAACAAT--TACAATTACTATTATTTACAAATTACA 735

RESULT 8
AY192160
LOCUS      AY192160
DEFINITION Synthetic construct multi-copy enhancer promoter element, complete
sequence.
ACCESSION  AY192160
VERSION     AY192160.1
KEYWORDS    synthetic construct
SOURCE      synthetic construct
ORGANISM    other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 827)
AUTHORS     Yao,Q., Peng,R. and Xiong,A.
TITLE       An artificial promoter containing many copies of enhancers can
increase phytase expression in transgenic rape
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 827)
AUTHORS     Yao,Q., Peng,R. and Xiong,A.
Direct Submission
TITLE       Direct Submission
JOURNAL     Submitted (07-DEC-2002) Biotechnology Research Center, Shanghai
Academy of Agriculture Science, Beidi Road 2901, Shanghai, Shanghai
201106, China
FEATURES             Location/Qualifiers
     source           1..827
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
     promoter         1..827
                     /note="multi-copy enhancer promoter element"
     misc_feature     1..90
                     /note="contains 3 copies of 30bp element from glycine
promoter"
     misc_feature     91..160
                     /note="contains 3 copies of 22bp W1 seed-specific
cis-element from pea lectin promoter"
     misc_feature     161..376
                     /note="contains 3 copies of 68bp UAS from phaseolin
promoter"
     misc_feature     501..827
                     /note="contains CaMV mini promoter (-60) and TMV omega
enhancer"

ORIGIN
Query Match      47.7%; Score 62; DB 12; Length 827;
Best Local Similarity 88.5%; Pred. No. 0.00019;
Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

Qy 12 AGGAAGTTCATTTTCATTTGGAATGGACACGCTGTGTCATTTCTCAACAATTACCAACAAC 71
Db 639 AGGAAGTTCATTTTCATTTGGAATGGACACGCTAT-----TTTACAAACAATTACCAACAAC 693
Qy 72 AACAAACAACAACAACAATTATACAAATTACTATTATTTACAAATTACA 115
Db 694 AACAAACAACAACAACAAT--TACAATTACTATTATTTACAAATTACA 735

RESULT 8
AY192160
LOCUS      AY192160
DEFINITION Synthetic construct multi-copy enhancer promoter element, complete
sequence.
ACCESSION  AY192160
VERSION     AY192160.1
KEYWORDS    synthetic construct
SOURCE      synthetic construct
ORGANISM    other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 827)
AUTHORS     Yao,Q., Peng,R. and Xiong,A.
TITLE       An artificial promoter containing many copies of enhancers can
increase phytase expression in transgenic rape
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 827)
AUTHORS     Yao,Q., Peng,R. and Xiong,A.
Direct Submission
TITLE       Direct Submission
JOURNAL     Submitted (07-DEC-2002) Biotechnology Research Center, Shanghai
Academy of Agriculture Science, Beidi Road 2901, Shanghai, Shanghai
201106, China
FEATURES             Location/Qualifiers
     source           1..827
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
     promoter         1..827
                     /note="multi-copy enhancer promoter element"
     misc_feature     1..90
                     /note="contains 3 copies of 30bp element from glycine
promoter"
     misc_feature     91..160
                     /note="contains 3 copies of 22bp W1 seed-specific
cis-element from pea lectin promoter"
     misc_feature     161..376
                     /note="contains 3 copies of 68bp UAS from phaseolin
promoter"
     misc_feature     501..827
                     /note="contains CaMV mini promoter (-60) and TMV omega
enhancer"

ORIGIN
Query Match      47.7%; Score 62; DB 12; Length 827;
Best Local Similarity 88.5%; Pred. No. 0.00019;
Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

Qy 12 AGGAAGTTCATTTTCATTTGGAATGGACACGCTGTGTCATTTCTCAACAATTACCAACAAC 71
Db 639 AGGAAGTTCATTTTCATTTGGAATGGACACGCTAT-----TTTACAAACAATTACCAACAAC 693
Qy 72 AACAAACAACAACAACAATTATACAAATTACTATTATTTACAAATTACA 115
Db 694 AACAAACAACAACAACAAT--TACAATTACTATTATTTACAAATTACA 735
```



ADQOVATCGTSGTLQVNVPHLMQPGEGYLYELCVAKTSQTECDIYPLRGIRSVAV  
KGEQFLNHKPFYFTGFRHEDADLRGKGFDMVMDHMDWIGANSIRTSHPYA  
EEMLDWADSHGIVVIDETAAGVNSLSLGI GFAGNKPKELYSEEA VNGETQQAHLQAI  
KELIARDKHGIVVIDETAAGVNSLSLGI GFAGNKPKELYSEEA VNGETQQAHLQAI  
AHTDTSIDLFDVLCNLRYYGMYVQSGDLETAEKVLEKELLAWQEKLHQPILII TEYGVD  
TLAGLHSMYTDWMSSEYOCALMDMTHRVFDRVS AVVGEQVWNFADFATSQGLRVGGN  
KKGIFTRDRKPKSAAPLLQKRWTMGNFGEKPOQGGKQ

misc\_feature

14114..15961

/note="similar to his tagged Gusa"

## ORIGIN

Query Match 47.7%; Score 62; DB 12; Length 16368;

Best Local Similarity 88.5%; Pred. No. 0.00012;

Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAATGGACACGTGTGTCATTCTCAACAATTACCAACAAC 71

Db 11351 AGGAAGTTCATTTCATTTGGAATGGACACGTGTGTCATTCTCAACAATTACCAACAAC 11297

QY 72 AACAAACAACAACAATATACAAATTACTATTACAAATTACA 115

Db 11296 AACAAACAACAACAACAT--TACAATTACTATTACAAATTACA 11255

## RESULT 11

AY178048/c 16520 bp DNA circular SYN 06-APR-2003

LOCUS Expression vector pYPX145, complete sequence.

ACCESSION AY178048

VERSION AY178048.1 GI:29568874

KEYWORDS Expression vector pYPX145

SOURCE Expression vector pYPX145

ORGANISM other sequences; artificial sequences; vectors.

REFERENCE 1 (bases 1 to 16520)

AUTHORS Yao Q., Peng R. and Xiong A.

TITLE Direct Submission

JOURNAL Submitted (12-NOV-2002) Shanghai Academy of Agricultural Sciences,

Agro-Biotechnology Research Center, 2901 Beidi Rd, Shanghai 201106,

China

COMMENT NCBI staff are still waiting for submitters to provide appropriate

coding region information.

FEATURES

Location/Qualifiers

1..16520

/organism="Expression vector pYPX145"

/mol\_type="genomic DNA"

/db\_xref="taxon:218200"

/lab\_host="Agrobacterium tumefaciens"

/gene="cat"

complement(5139..5798)

complement(5139..5798)

/gene="cat"

/codon\_start=1

/product="chloramphenicol acetyl transferase"

/protein\_id="AA084029.1"

/db\_xref="GI:29568875"

/translation="MEKKITGYTTVDISOMHKEHFEAPQSAQCTYNTQVOLDITAF

LTKVKNKHFPYFIHLRLNNAHPEFRAMKDGELVINDSVHPCTYVPEQTEF

SSLSVEYHDDFQFLHYISQDVACGENLAYPPKGFIEFNFFVSANPWVSTFSLNV

ANMDFAPVFTMGKYYTQGDVKVLPALQVHHAVDCGFHVGRLNELQYCDVQGG

A"

complement(6529..7504)

/note="similar to aminoglycoside-3'-O-phosphotransferase"

complement(9589..11400)

/gene="gus"

complement(9589..11400)

/gene="gus"

/codon\_start=1

/product="Gus"

/protein\_id="AA084030.1"

/db\_xref="GI:29568876"

/translation="MLRPVETPTREIKKLDGLWAFSLDRENGCIDQRWWSALQESRA

IAPGSDNQPADADIRNAYGNVYQREVFIPKGWAGQRIVLRFDAVTHYGVWYNNQ

EMVEHQGYTPFEADYTPYVIAGKSVRLTVCVNNELNHQTIPPGMVIITDENGKKOSY  
FHDFFNYAGIHRSSVMLYTPNTWDDITVTHVAQDCNHSVDMQVANGDVSELRD  
ADQOVATCGTSGTLQVNVPHLMQPGEGYLYELCVAKTSQTECDIYPLRGIRSVAV  
KGEQFLNHKPFYFTGFRHEDADLRGKGFDMVMDHMDWIGANSIRTSHPYA  
EEMLDWADSHGIVVIDETAAGVNSLSLGI GFAGNKPKELYSEEA VNGETQQAHLQAI  
KELIARDKHGIVVIDETAAGVNSLSLGI GFAGNKPKELYSEEA VNGETQQAHLQAI  
AHTDTSIDLFDVLCNLRYYGMYVQSGDLETAEKVLEKELLAWQEKLHQPILII TEYGVD  
TLAGLHSMYTDWMSSEYOCALMDMTHRVFDRVS AVVGEQVWNFADFATSQGLRVGGN  
KKGIFTRDRKPKSAAPLLQKRWTMGNFGEKPOQGGKQ

misc\_feature

14251..16113

/note="similar to his tagged Gusa"

## ORIGIN

Query Match 47.7%; Score 62; DB 12; Length 16520;

Best Local Similarity 88.5%; Pred. No. 0.00012;

Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAATGGACACGTGTGTCATTCTCAACAATTACCAACAAC 71

Db 11503 AGGAAGTTCATTTCATTTGGAATGGACACGTGTGTCATTCTCAACAATTACCAACAAC 11449

QY 72 AACAAACAACAACAATATACAAATTACTATTACAAATTACA 115

Db 11448 AACAAACAACAACAACAT--TACAATTACTATTACAAATTACA 11407

## RESULT 12

A94910 1006 bp DNA linear PAT 26-JAN-2000

LOCUS Sequence 22 from Patent WO9931258.

ACCESSION A94910

VERSION A94910.1 GI:6779107

KEYWORDS unidentified

SOURCE unidentified

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 1006)

AUTHORS Stuiver, M.H. and Sibbols, F.H.

TITLE CONSTITUTIVE PLANT PROMOTERS

JOURNAL Patent: WO 9931258-A 22 24-JUN-1999;

STUIVER MAARTEN HENDRIK (NL); MOGEN INT (NL)

FEATURES

Location/Qualifiers

1..1006

/organism="unidentified"

/mol\_type="unassigned DNA"

/db\_xref="taxon:32644"

## ORIGIN

Query Match 46.9%; Score 61; DB 6; Length 1006;

Best Local Similarity 88.3%; Pred. No. 0.00031;

Matches 91; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAATGGACACGTGTGTCATTCTCAACAATTACCAACAAC 71

Db 906 AGGAAGTTCATTTCATTTGGAATGGACACGTGTGTCATTCTCAACAATTACCAACAAC 960

QY 72 AACAAACAACAACAATATACAAATTACTATTACAAATTACA 114

Db 961 AACAAACAACAACAACAT--TACAATTACTATTACAAATTAC 1001

## RESULT 13

BD134819 1006 bp DNA linear PAT 18-SEP-2002

LOCUS Constitutive plant promoter.

ACCESSION BD134819

VERSION BD134819.1 GI:23229764

KEYWORDS JP 2002508186-A/22.

SOURCE unidentified

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 1006)

AUTHORS Stuiver, M.H. and Sibbols, F.H.

```
TITLE      Constitutive plant promoter
JOURNAL    Patent: JP 2002508186-A 22-19-MAR-2002;
COMMENT    MOGEN INTERNATIONAL NV
           OS Unidentified
           PN JP 2002508186-A/22
           PD 19-MAR-2002
           PF 10-DEC-1998 JP 2000539156
           PR 12-DEC-1997 EP 97203912.7
           PI MAARTEN HENDRIK STUIVER, FLOOR HENDRIC SIBORTZ PC
           C12N15/09, C12N15/00
           CC Strandedness: Double;
           CC Topology: Linear;
           CC Constitutive plant promoter
           FH Key Location/Qualifiers
           FT source 1..1006
           FT Location/Qualifiers
           source 1..1006
           /organism="unidentified"
           /mol_type="genomic DNA"
           /db_xref="taxon:32644"

ORIGIN
Query Match 46.9%; Score 61; DB 6; Length 1006;
Best Local Similarity 88.3%; Pred. No. 0.00031;
Matches 91; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTGGAATGGACACGTGTCATTCTCAACAATTACCAACAAC 71
Db 906 AGGAAGTTCATTTCATTGAGAGGACAGTAT-----TTTACACAATTACCAACAAC 960

QY 72 AACAAACAACAACAATTATACAAATTACTATTACAATTAC 114
Db 961 AACAAACAACAACAACAT--TACAATTACTATTACAATTAC 1001

RESULT 14
A59870 LOCUS 1138 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 2 from Patent WO9706269.
ACCESSION A59870
VERSION A59870.1 GI:3715061
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1
AUTHORS Jepson, I.
TITLE INDUCIBLE HERBICIDE RESISTANCE
JOURNAL Patent: WO 9706269-A 2 20-FEB-1997;
ZENECA LTD (GB)
FEATURES
source 1..1138
Location/Qualifiers
1..1138
/organism="unidentified"
/mol_type="unassigned DNA"
/isolate="PLASMID MJB1"
/db_xref="taxon:32644"

ORIGIN
Query Match 46.5%; Score 60.4; DB 6; Length 1138;
Best Local Similarity 85.8%; Pred. No. 0.00041;
Matches 91; Conservative 0; Mismatches 11; Indels 4; Gaps 2;

QY 12 AGGAAGTTCATTTCATTGGAATGGACACGTGTCATTCTCAACAATTACCAACAAC 71
Db 743 AGGAAGTTCATTTCATTGGAGAGG--ACCTCGAGTATTTTACACAATTACCAACAAC 800

QY 72 AACAAACAACAACAATTATACAAATTACTATTACAATTACATC 117
Db 801 AACAAACAACAACAACAT--TACAATTACTATTACAATTACACC 844

Search completed: July 3, 2005, 03:19:09
Job time : 724.588 secs
```

RESULT 15

the ruge blark (uspio)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 02:18:08 ; Search time 140.471 Seconds  
(without alignments)  
3634.343 Million cell updates/sec

Title: US-10-814-858A-1  
Perfect score: 312  
Sequence: 1 gtcgaccatcattgaagg.....tgacgcacttgacgtact 312

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/prodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	310.4	99.5	452	4	US-09-263-692A-1
2	46	14.7	46	4	US-09-263-692A-5
3	46	14.7	46	4	US-09-263-692A-6
c 4	41.6	13.3	7218	1	US-08-232-463-14
5	36.8	11.8	254778	4	US-09-949-016-12417
6	36.4	11.7	289	3	US-09-007-005-17
7	36.4	11.7	289	3	US-09-244-796-17
c 8	33.8	10.8	22294	4	US-09-949-016-14020
9	33.6	10.8	248	3	US-09-007-005-32
10	33.6	10.8	248	3	US-09-244-796-32
11	33.6	10.8	277	3	US-09-007-005-3
12	33.6	10.8	277	3	US-09-244-796-3
13	33.4	10.7	601	4	US-09-949-016-78043
c 14	32.4	10.4	1096	3	US-08-858-207A-136
c 15	32.4	10.4	11303	3	US-08-861-527-115
c 16	32.4	10.4	68702	4	US-09-949-016-16328
17	32.4	10.4	102409	4	US-09-949-016-15148
18	32.2	10.3	3270	4	US-09-248-796A-4969
19	31.6	10.1	1958	4	US-09-270-767-10139
20	31.6	10.1	2526	3	US-09-202-712-1
21	31.6	10.1	84495	3	US-09-797-906-3
c 22	31.2	10.0	491	4	US-09-270-767-7112
c 23	31.2	10.0	491	4	US-09-270-767-22394
c 24	31	9.9	601	4	US-09-949-016-164450
c 25	31	9.9	44715	4	US-09-949-016-16353
26	30.8	9.9	1980	4	US-09-583-110-1382
27	30.8	9.9	1986	4	US-09-107-433-1097

28	30.6	9.8	5053	3	US-08-961-527-187	Sequence 187, Appl
c 29	30.6	9.8	74962	4	US-09-685-853A-3	Sequence 3, Appl1
30	30.4	9.7	207	4	US-09-248-796A-6311	Sequence 6311, Ap
c 31	30.4	9.7	89689	4	US-09-949-016-13089	Sequence 13089, A
c 32	30.4	9.7	153866	4	US-09-949-016-16919	Sequence 16919, A
33	30.2	9.7	477	4	US-09-669-751-206	Sequence 206, App
c 34	30.2	9.7	1069	4	US-09-270-767-13807	Sequence 13807, A
c 35	30.2	9.7	177669	4	US-09-949-016-13713	Sequence 13713, A
36	30	9.6	1214	4	US-09-506-286B-15	Sequence 15, Appl
37	30	9.6	1214	4	US-09-506-286B-18	Sequence 18, Appl
38	30	9.6	1214	4	US-09-762-861B-15	Sequence 15, Appl
39	30	9.6	1214	4	US-09-762-861B-18	Sequence 18, Appl
40	30	9.6	1214	4	US-10-065-133A-15	Sequence 15, Appl
41	30	9.6	1214	4	US-10-065-133A-18	Sequence 18, Appl
42	30	9.6	1214	4	US-10-434-811A-15	Sequence 15, Appl
43	30	9.6	1214	4	US-10-434-811A-18	Sequence 18, Appl
44	30	9.6	1241	4	US-09-506-286B-13	Sequence 13, Appl
45	30	9.6	1241	4	US-09-506-286B-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-263-692A-1  
; Sequence 1, Application US/09263692A  
; Patent No. 6639065  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A chemically synthesized artificial promoter for high level expr  
; TITLE OF INVENTION: transgenes and a method for its synthesis  
; FILE REFERENCE: Q52511  
; CURRENT APPLICATION NUMBER: US/09/263,692A  
; CURRENT FILING DATE: 1999-03-05  
; PRIOR APPLICATION NUMBER: 3322/Del/98  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 452  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: synthetic DNA promoter sequence  
US-09-263-692A-1

Query Match	99.5%	Score	310.4	DB	4	Length	452
Best Local Similarity	99.7%	Pred. No.	1.7e-94				
Matches	311	Conservative	0	Mismatches	1	Indels	0
Gaps	0						
QY	1	GTCGACCATCATTGCAAGGGCCCTCGGTAATACCATTTGGGAAAAGTTGGTAATACGGA	60				
Db	1	GTCGACCATCATTGCAAGGGCCCTCGGTAATACCATTTGGGAAAAGTTGGTAATACGGA	60				
QY	61	AAAAGAAGATTTCATCCAGAAAAGTGTGAAAAAGTTGCGATTGCGTGGAAAAAGTTT	120				
Db	61	AAAAGAAGATTTCATCCAGAAAAGTGTGAAAAAGTTGCGATTGCGTGGAAAAAGTTT	120				
QY	121	CGATCTGACCATCTCTAGATCGTGAAAAGTTCAGTTAGCGCTTACGTACATATGTGG	180				
Db	121	CGATCTGACCATCTCTAGATCGTGAAAAGTTCAGTTAGCGCTTACGTACATATGTGG	180				
QY	181	ATTGTGGAAAAGAGACGAGGCGATCGTGGAAAAGAGCTTGTACGCTGTACGCTGA	240				
Db	181	ATTGTGGAAAAGAGAGCGGCGATCGTGGAAAAGAGCTTGTACGCTGTACGCTGA	240				
QY	241	CGATAGATAGATACACGTGCGACCGCTCCACTTTGACGCAATTTGACGCAATGACGCA	300				
Db	241	CGATAGATAGATACACGTGCGACCGCTCCACTTTGACGCAATTTGACGCAATGACGCA	300				
QY	301	CTTGACGCTACT	312				
Db	301	CTTGACGCTACT	312				

```
RESULT 2
US-09-263-692A-5
; Sequence 5, Application US/09263692A
; Patent No. 6639065
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A chemically synthesized artificial promoter for high level expression of transgenes and a method for its synthesis
; FILE REFERENCE: 052511
; CURRENT APPLICATION NUMBER: US/09/263,692A
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: 3322/Del/98
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 46
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: minimal domain (b)
US-09-263-692A-5

Query Match      14.7%; Score 46; DB 4; Length 46;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      267 CCACCTTGACGCAACAATTGACGCAACAATGACGCCCACTTGACGCTACT 312
Db      1 CCACCTTGACGCAACAATTGACGCAACAATGACGCCCACTTGACGCTACT 46

RESULT 3
US-09-263-692A-6
; Sequence 6, Application US/09263692A
; Patent No. 6639065
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A chemically synthesized artificial promoter for high level expression of transgenes and a method for its synthesis
; FILE REFERENCE: 052511
; CURRENT APPLICATION NUMBER: US/09/263,692A
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: 3322/Del/98
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 46
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: domain I
US-09-263-692A-6

Query Match      14.7%; Score 46; DB 4; Length 46;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      221 GCTTGTACGCTGTACGCTGACGATAGATAGATACACGTGCACCGT 266
Db      1 GCTTGTACGCTGTACGCTGACGATAGATAGATAGATACACGTGCACCGT 46

RESULT 4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
```

```
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: pTZgpt-Fls
; IMMEDIATE SOURCE:
US-08-232-463-14

Query Match      13.3%; Score 41.6; DB 1; Length 7218;
Best Local Similarity 4.2%; Pred. No. 0.0025;
Matches 11; Conservative 152; Mismatches 101; Indels 0; Gaps 0;

QY      15 GAAAGGGCCTCGGTAATACCATTTGTGGAAGAAAGTTGGTAATACGGAAGAAAGATTTCAT 74
Db      1307 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1248

QY      75 CATCCAGAAAGGTGTGGAAGAAAGTTGTGGATTGCGTGGAAAGTTCGATCTGACCATCT 134
Db      1247 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1188

QY      135 CTAGATCGTGGAAAGTTACGTTAGCGCTTACGATATGTGATTGTGGAAGAAAGA 194
Db      1187 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1128

QY      195 AGACGAGGCGATCGGTGGAAGAAAGCTGTGACGCTGACGTCACGATAGATAC 254
Db      1127 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1068

QY      255 ACGTGCACGCTCCACTTGACGCA 278
Db      1067 RATCGCAAGCTCCCTCGACCTGCA 1044

RESULT 5
US-09-949-016-12417
; Sequence 12417, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```



```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14020
; LENGTH: 22294
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-14020
```

```
Query Match 10.8%; Score 33.8; DB 4; Length 22294;
Best Local Similarity 53.4%; Pred. No. 1.8;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 123 ATCTGACCACTCTTAGATCGTGGAAAAAGTTACGTTAGCGCTTACGTACATATGGGAT 182
Db 4766 AACTGTATTGTCAATGTCTTAAATGCGAAAAATTCATAGATTACATAACATGGAGAA 4707

QY 183 TGTGAAAAGAGACGGAGGATCGTGTGAAAAAGAGCTTGTACGCTGTACGCTGACG 242
Db 4706 TGAGGAAAATATAAAACAAGTAATGATGAAAAATGATGTATCATTTAAAAAGCGGTG 4647

QY 243 ATAGATAGATACA 255
Db 4646 ATATAAAATTAA 4634
```

```
RESULT 9
US-09-007-005-32
; Sequence 32, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 248
; TYPE: RNA
; ORGANISM: Homo sapiens
; US-09-007-005-32
```

```
Query Match 10.8%; Score 33.6; DB 3; Length 248;
Best Local Similarity 18.2%; Pred. No. 0.23;
Matches 28; Conservative 65; Mismatches 61; Indels 0; Gaps 0;

QY 136 TAGATCGTGGAAAAAGTTACGTTAGCGCTTACGTACATATGGGATTTGGAAAAAGAA 195
Db 62 URGARARGARARCRGRARARCRURGRARURCRURGRARARCRGRARARGRAR 121

QY 196 GACGGAGGATCGTGGAAAAAGAGCTTGTACGCTGTACGCTGACGATAGATACATA 255
```

```
Db 122 CCRURGRURGRCRGRURARARARCRURCRURGRARARCRARCRURGRARAR 181

QY 256 COTGACGCGTCCACTTGAACGACCAATTTGACGCA 289
Db 182 APCRARCRARARARCRURGRGRARARCRARGCR 215

RESULT 10
US-09-244-796-32
; Sequence 32, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 248
; TYPE: RNA
; ORGANISM: Homo sapiens
; US-09-244-796-32
```

```
Query Match 10.8%; Score 33.6; DB 3; Length 248;
Best Local Similarity 18.2%; Pred. No. 0.23;
Matches 28; Conservative 65; Mismatches 61; Indels 0; Gaps 0;

QY 136 TAGATCGTGGAAAAAGTTACGTTAGCGCTTACGTACATATGTGATTTGGAAAAAGAA 195
Db 62 URGARARGARARCRGRARARCRURGRARURCRURGRARARCRURGRARARGRAR 121

QY 196 GACGGAGGATCGTGGAAAAAGAGCTTGTACGCTGTACGCTGACGATAGATACATA 255
Db 122 CCRURGRURGRCRGRURARARARCRURCRURGRARARCRARCRURGRARAR 181

QY 256 COTGACGCGTCCACTTGAACGACCAATTTGACGCA 289
Db 182 APCRARCRARARARCRURGRGRARARCRARGCR 215
```

```
RESULT 11
US-09-007-005-3
; Sequence 3, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 277
; TYPE: RNA
```



TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:

INFORMATION FOR SEQ ID NO: 136:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1096 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 858-207A-136

Query Match 10.4%; Score 32.4; DB 3; Length 1096;

Matches	75	Conservative	0	Mismatches	71	Indels	0	Gaps	0
62	AAAGAAGATT	CATCAT	CAGAAAAAGGT	TGGAAAAAGTT	TGTGGATT	TGCGTGGAAAAAGTTC	121		
699	AGACAAGGTAA	TTTACT	TTTGAAGAAGAT	GGTTCAGAGTTT	TATCTGTTGTT	GTGAATCAATGA	640		
122	GATCTGACCAT	CTCTAGAT	CTGTGGAAAAAGTT	TCAGTTAGCGCTT	ACGTACATAT	TGTGGA	181		
639	GTTTGAACAA	TCTTTT	TATGGAATTTAA	CACGCGGAAGAA	TTTAAGTTCAGTT	GTGTTTT	580		
182	TTGTGAAAAA	AGAACG	GAGGCATC	207					
579	TTCTGAAAAA	TTGCAAGT	TTCCGCATC	554					

RESULT 15  
 US-08-961-527-115/c  
 ; Sequence 115, Application US/08961527  
 ; Patent No. 6420135  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 391  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/961,527  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brookes, A. Anders  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340PI  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 115:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 11303 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-961-527-115

Query Match	10.4%	Score 32.4;	DB 3;	Length 11303;
Best Local Similarity	51.4%	Pred. No. 3.9;		
Matches 75: Conservative	0;	Mismatches 71;	Indels 0;	Gaps 0;

Qy	62	AAAGAAGTTCATC	ATCCAGAAAGCTG	CGAAGCTGTG	GGATTCGTCG	CGAAGAGTTC	121
Db	6373	AGACAGGTAAAT	TCTTGGAAAGAT	TGGTTAGAGAT	TATCTGTG	TGTAACAATGA	6314
Qy	122	GATCTGACCATCT	CTAGATCGTGG	AAAAAGTTCAC	GTTCAGGCTT	ACGTACATATG	181
Db	6313	GTTTGACAACTC	TCTTACTTTAT	TGGAATTAACA	GCGCGAAGAAT	TAAAGTTCAGTTG	6254
Qy	182	TTGTGAAAAA	GAGACG	GAGGCATC	207		
Db	6253	TTCTGAAAAA	ATTGCG	AGTTC	CCGCATC	6228	

Search completed: July 3, 2005, 04:55:25  
Job time : 143.471 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 00:55:42 ; Search time 1729.41 Seconds  
(without alignments)  
8741.717 Million cell updates/sec

Title: US-10-814-858A-1

Perfect score: 312  
Sequence: 1 gtcgaccatcattgaaagg.....tgacgcacttgacgtact 312

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 segs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_cm.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	310.4	99.5	452	6	E64760	E64760 Chemically
2	310.4	99.5	452	6	AR427842	AR427842 Sequence
3	46	14.7	46	6	E64764	E64764 Chemically
4	46	14.7	46	6	E64765	E64765 Chemically
5	46	14.7	46	6	AR427845	AR427845 Sequence
6	46	14.7	46	6	AR427846	AR427846 Sequence
7	41.6	13.3	7218	6	I66494	I66494 Sequence 14
8	41	13.1	114519	8	AC124964	AC124964 Medicago
9	41	13.1	121565	2	AC135605	AC135605 Medicago
10	40.6	13.0	172623	2	AC147540	AC147540 Pan trogl
11	40.6	13.0	197004	2	AC145064	AC145064 Pan trogl
12	39.2	12.6	134442	8	AP004619	AP004619 Oryza sat
13	39.2	12.6	150300	8	AP003878	AP003878 Oryza sat
14	39.2	12.6	182932	2	AC112060	AC112060 Rattus no
15	39.2	12.6	185618	2	AC108612	AC108612 Rattus no
16	38.8	12.4	330362	2	AC129990	AC129990 Rattus no
17	36.8	11.8	63114	5	CR377227	CR377227 Zebraphish
18	36.8	11.8	139006	9	AL162431	AL162431 Human DNA
19	36.8	11.8	148068	9	CNS01RHN	AL161871 Human chr

20	36.4	11.7	289	6	AR162089	AR162089 Sequence
21	36.4	11.7	289	6	AR166614	AR166614 Sequence
22	35.8	11.5	177607	2	AC131994	AC131994 Mus muscu
23	35.8	11.5	191924	2	AC148087	AC148087 Mus muscu
24	35.6	11.4	110000	2	AC146850_2	Continuation (3 of
25	35.4	11.3	90902	8	ATF28M20	AL031004 Arabidops
26	35.4	11.3	102495	3	CEY80D3A	AL132853 Caenorhab
27	35.4	11.3	111517	8	ATF11C18	AL049607 Arabidops
28	35.4	11.3	147891	5	BX255947	BX255947 Zebraphish
29	35.4	11.3	151551	2	AC021291	AC021291 Homo sapi
30	35.4	11.3	179187	9	AC091549	AC091549 Homo sapi
31	35.4	11.3	192053	9	AC090155	AC090155 Homo sapi
32	35.4	11.3	196242	8	ATCHRIV75	AL161579 Arabidops
33	35.4	11.3	221126	2	AC021184	AC021184 Homo sapi
34	35.2	11.3	25593	3	CEH21P03	Z95619 Caenorhabdi
35	35.2	11.3	49261	2	AC017912	AC017912 Drosophil
36	35.2	11.3	89370	1	AE015945	AE015945 Clostridi
37	35.2	11.3	176329	2	AC117968	AC117968 Rattus no
38	35.2	11.3	176510	3	AC007054	AC007054 Drosophil
39	35.2	11.3	225790	2	AC099213	AC099213 Rattus no
40	35.2	11.3	226637	2	AC095337	AC095337 Rattus no
41	35.2	11.3	245875	2	AC125836	AC125836 Rattus no
42	35.2	11.3	301129	3	AE003785	AE003785 Drosophil
43	35.2	11.3	318312	2	AC097988	AC097988 Rattus no
44	35	11.2	113451	2	AC023297	AC023297 Homo sapi
45	35	11.2	139315	2	BX537131	BX537131 Danio rer

## ALIGNMENTS

### RESULT 1

E64760 452 bp DNA linear PAT 31-JAN-2002  
LOCUS Chemically synthesized artificial promoter for realizing high-level  
DEFINITION expression of introduced gene and method for synthesizing the  
same.

ACCESSION E64760

VERSION E64760.1 GI:18628523

KEYWORDS JP 2000139477-A/1.

SOURCE synthetic construct

ORGANISM other sequences; artificial sequences.

REFERENCE 1 (bases 1 to 452)

AUTHORS Tori, R., Sawanto, S.B., Synge, P.K. and Gupta, S.K.

TITLE Chemically synthesized artificial promoter for realizing high-level

expression of introduced gene and method for synthesizing the same

JOURNAL Patent: JP 2000139477-A 1 23-MAY-2000;

COMMENT COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH

OS Artificial Sequence

PN JP 2000139477-A/1

PD 23-MAY-2000

PF 27-APR-1999 JP 1999119227

PP 09-NOV-1998 IN 3322/98

PI RAKESH TORI, SALLY BISHWANATO SAWANTO, PURAJUNNA KUMAR SYNGE,

PC SHIFU KUMAR GUPTA

CC C12N15/09, C12N5/10, C12N15/00, C12N5/00

FT Key Location/Qualifiers

FT source 1.452 /organism='Artificial Sequence'

FT Location/Qualifiers

1.452 /organism='synthetic construct'

/mol\_type='genomic DNA'

/db\_xref='taxon:32630'

### ORIGIN

Query Match 99.5%; Score 310.4; DB 6; Length 452;

Best Local Similarity 99.7%; Pred. No. 6.4e-79;

Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACCATCATTTGAAAGGGCTCGGTAATACCATTTGGGAAAGTTGGTAATACGGA 60

|||||  
Db 1 GTGCACCATCTTTGAAAGGCGCTCGGTATACCATTTGGAAAAAGTTGGTAAATACGGA 60  
QY 61 AAAAGAAGATTTCATCATCCAGAAAAAGTGTGGAAAAAGTTGTGGATTGCGTGGAAAAAGTT 120  
Db 61 AAAAGAAGATTTCATCATCCAGAAAAAGTGTGGAAAAAGTTGTGGATTGCGTGGAAAAAGTT 120  
QY 121 CGATCTGACCATCTCTAGATCGTGGAAAAAGTTACGCTTAGCGCTTACGTACATATGTGG 180  
Db 121 CGATCTGACCATCTCTAGATCGTGGAAAAAGTTACGCTTAGCGCTTACGTACATATGTGG 180  
QY 181 ATTGTGAAAAAGAACGAGCGGCATCGTGGAAAAAGTGTGACGCTGTACGCTGACGCTGA 240  
Db 181 ATTGTGAAAAAGAACGAGCGGCATCGTGGAAAAAGTGTGACGCTGTACGCTGTACGCTGA 240  
QY 241 CGATAGATAGATACAGTGCACGCGTCCACTTGACGCACAATTTGACGCAATGACGCCA 300  
Db 241 CGATAGATAGATACAGTGCACGCGTCCACTTGACGCACAATTTGACGCAATGACGCCA 300  
QY 301 CTTGACGCTACT 312  
Db 301 CTTGACGCTACT 312

RESULT 2  
AR427842  
LOCUS 452 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 1 from patent US 6639065.  
ACCESSION AR427842  
VERSION AR427842.1 GI:40186826  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 452)  
AUTHORS Tuli, R., Sawant, S.V., Singh, P.K. and Gupta, S.K.  
TITLE Chemically synthesized artificial promoter for high level expression of transgenes  
JOURNAL Patent: US 6639065-A 1 28-OCT-2003;  
FEATURES  
source Location/Qualifiers  
1..452  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 99.5%; Score 310.4; DB 6; Length 452;  
Best Local Similarity 99.7%; Pred. No. 6.4e-79;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGCACCATCTTTGAAAGGCGCTCGGTATACCATTTGGAAAAAGTTGGTAAATACGGA 60  
Db 1 GTGCACCATCTTTGAAAGGCGCTCGGTATACCATTTGGAAAAAGTTGGTAAATACGGA 60  
QY 61 AAAAGAAGATTTCATCATCCAGAAAAAGTGTGGAAAAAGTTGTGGATTGCGTGGAAAAAGTT 120  
Db 61 AAAAGAAGATTTCATCATCCAGAAAAAGTGTGGAAAAAGTTGTGGATTGCGTGGAAAAAGTT 120  
QY 121 CGATCTGACCATCTCTAGATCGTGGAAAAAGTTACGCTTAGCGCTTACGTACATATGTGG 180  
Db 121 CGATCTGACCATCTCTAGATCGTGGAAAAAGTTACGCTTAGCGCTTACGTACATATGTGG 180  
QY 181 ATTGTGAAAAAGAACGAGCGGCATCGTGGAAAAAGTGTGACGCTGTACGCTGACGCTGA 240  
Db 181 ATTGTGAAAAAGAACGAGCGGCATCGTGGAAAAAGTGTGACGCTGTACGCTGTACGCTGA 240  
QY 241 CGATAGATAGATACAGTGCACGCGTCCACTTGACGCACAATTTGACGCAATGACGCCA 300  
Db 241 CGATAGATAGATACAGTGCACGCGTCCACTTGACGCACAATTTGACGCAATGACGCCA 300  
QY 301 CTTGACGCTACT 312  
Db 301 CTTGACGCTACT 312

RESULT 3  
E64764  
LOCUS 46 bp DNA linear PAT 31-JAN-2002  
DEFINITION Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same.

ACCESSION E64764  
VERSION E64764.1 GI:18628527  
KEYWORDS JP 2000139477-A/5.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 46)  
AUTHORS Tori, R., Sawanto, S.B., Synge, P.K. and Guputa, S.K.  
TITLE Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same  
JOURNAL Patent: JP 2000139477-A 5 23-MAY-2000;  
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH

COMMENT OS Artificial Sequence  
PN JP 2000139477-A/5  
PD 23-MAY-2000  
PF 27-APR-1999 JP 1999119227  
PI 09-NOV-1998 IN 3322/98  
PI RAKESH TORI, SALLY BISHWANATO SAWANTO, PURAJUNNA KUMAR SYNGE,  
PI SHIFU KUMAR GUPUTA  
PC C12N15/09, C12N5/10, C12N15/00, C12N5/00  
CC  
FH Key Location/Qualifiers  
FT source 1..46  
/organism='Artificial Sequence'.  
FEATURES  
source Location/Qualifiers  
1..46  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"

ORIGIN  
Query Match 14.7%; Score 46; DB 6; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.02;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 267 CCACCTTGACGCAATTTGACGCACAAATGACGCCCAATGACGCTACT 312  
Db 1 CCACCTTGACGCAATTTGACGCACAAATGACGCCCAATGACGCTACT 46

RESULT 4  
E64765  
LOCUS 46 bp DNA linear PAT 31-JAN-2002  
DEFINITION Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same.

ACCESSION E64765  
VERSION E64765.1 GI:18628528  
KEYWORDS JP 2000139477-A/6.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 46)  
AUTHORS Tori, R., Sawanto, S.B., Synge, P.K. and Guputa, S.K.  
TITLE Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same  
JOURNAL Patent: JP 2000139477-A 6 23-MAY-2000;  
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH

COMMENT OS Artificial Sequence  
PN JP 2000139477-A/6  
PD 23-MAY-2000  
PF 27-APR-1999 JP 1999119227  
PI 09-NOV-1998 IN 3322/98  
PI RAKESH TORI, SALLY BISHWANATO SAWANTO, PURAJUNNA KUMAR SYNGE,  
PI SHIFU KUMAR GUPUTA  
PC C12N15/09, C12N5/10, C12N15/00, C12N5/00



```
CC Key Location/Qualifiers
FH source 1..46
FT /organism='Artificial Sequence'.
FT Location/Qualifiers
FEATURES
  source 1..46
    /organism="synthetic construct"
    /mol_type="genomic DNA"
    /db_xref="taxon:32630"
ORIGIN
  Query Match 14.7%; Score 46; DB 6; Length 46;
  Best Local Similarity 100.0%; Pred. No. 0.02;
  Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  221 GCTTGTACGCTGACGCTGACGATAGATAGATACACGTCGACGCGT 266
  Db 1 GCTTGTACGCTGACGCTGACGATAGATAGATACACGTCGACGCGT 46
RESULT 7
LOCUS I66494/c
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
  source 1..7218
    Location/Qualifiers
    /organism="unknown"
    /mol_type="unassigned DNA"
ORIGIN
  Query Match 13.3%; Score 41.6; DB 6; Length 7218;
  Best Local Similarity 4.2%; Pred. No. 0.52;
  Matches 11; Conservative 152; Mismatches 101; Indels 0; Gaps 0;
  15 GAAAGCGCTCGGTATACCATTTGTGAAAAAGTTGGTAATACGAAAAAGAAATTTCAT 74
  Db 1307 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1248
  QY 75 CATCCAGAAAAGGTGCGAAAAGTTGCGATTGCGTGGAAAAGTTGCGATCTGACCATCT 134
  Db 1247 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1188
  QY 135 CTAGATCGTGGAAAAGTTACGTTAGCGCTTACGATATCTGCGATTGTGAAAAAGA 194
  Db 1187 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1128
  QY 195 AGACGAGGACATCGGTGGAAGAAGCTTGTACGCTGACGTCGACGATAGATAGATAC 254
  Db 1127 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1068
  QY 255 ACGTGACGCGTCCACTTGACGCA 278
  Db 1067 RATCGCAAGCTCCCTCGACCTGCA 1044
RESULT 8
LOCUS AC124964/c
DEFINITION Medicago truncatula clone mth2-27c4, complete sequence.
ACCESSION AC124964
VERSION AC124964.17 GI:45120181
KEYWORDS HTG.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
REFERENCE 1 (bases 1 to 114519)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
TITLE Cook, D., Kim, D. and Roe, B. A.
JOURNAL Medicago truncatula BAC Clone mth2-27c4
REFERENCE 2 (bases 1 to 114519)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
  Cook, D., Kim, D. and Roe, B. A.
```

```
CC Key Location/Qualifiers
FH source 1..46
FT /organism='Artificial Sequence'.
FT Location/Qualifiers
FEATURES
  source 1..46
    /organism="synthetic construct"
    /mol_type="genomic DNA"
    /db_xref="taxon:32630"
ORIGIN
  Query Match 14.7%; Score 46; DB 6; Length 46;
  Best Local Similarity 100.0%; Pred. No. 0.02;
  Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  221 GCTTGTACGCTGACGCTGACGATAGATAGATACACGTCGACGCGT 266
  Db 1 GCTTGTACGCTGACGCTGACGATAGATAGATACACGTCGACGCGT 46
RESULT 5
LOCUS AR427845
DEFINITION Sequence 5 from patent US 6639065.
ACCESSION AR427845
VERSION AR427845.1 GI:40186829
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 46)
AUTHORS Tuli, R., Sawant, S. V., Singh, P. K. and Gupta, S. K.
TITLE Chemically synthesized artificial promoter for high level
  expression of transgenes
JOURNAL Patent: US 6639065-A 5 28-OCT-2003;
FEATURES
  source 1..46
    Location/Qualifiers
    /organism="unknown"
    /mol_type="genomic DNA"
ORIGIN
  Query Match 14.7%; Score 46; DB 6; Length 46;
  Best Local Similarity 100.0%; Pred. No. 0.02;
  Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  267 CCACCTTGACGCAATGACGCAATGACGCAATGACGCAATGACGCTACT 312
  Db 1 CCACCTTGACGCAATGACGCAATGACGCAATGACGCAATGACGCTACT 46
RESULT 6
LOCUS AR427846
DEFINITION Sequence 6 from patent US 6639065.
ACCESSION AR427846
VERSION AR427846.1 GI:40186830
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 46)
AUTHORS Tuli, R., Sawant, S. V., Singh, P. K. and Gupta, S. K.
TITLE Chemically synthesized artificial promoter for high level
  expression of transgenes
JOURNAL Patent: US 6639065-A 6 28-OCT-2003;
FEATURES
  source 1..46
    Location/Qualifiers
    /organism="unknown"
    /mol_type="genomic DNA"
ORIGIN
  Query Match 14.7%; Score 46; DB 6; Length 46;
  Best Local Similarity 100.0%; Pred. No. 0.02;
```



REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 172623)  
Li,J. and Roe,B.A.  
Direct Submission  
Submitted (06-FEB-2004) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA

## COMMENT

On Feb 6, 2004 this sequence version replaced gi:39573794.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2143: contig of 2143 bp in length  
2144 2243: gap of unknown length  
2244 4771: contig of 2528 bp in length  
4772 4871: gap of unknown length  
4872 7178: contig of 2307 bp in length  
7179 7278: gap of unknown length  
7279 9321: contig of 2043 bp in length  
9322 9421: gap of unknown length  
9422 12171: contig of 2750 bp in length  
12172 12271: gap of unknown length  
12272 14443: contig of 2172 bp in length  
14444 14543: gap of unknown length  
14544 17736: contig of 3193 bp in length  
17737 17837: gap of unknown length  
17838 21350: contig of 3514 bp in length  
21351 21450: gap of unknown length  
21451 24909: contig of 3459 bp in length  
24910 25009: gap of unknown length  
25010 28464: contig of 3455 bp in length  
28465 28564: gap of unknown length  
28565 32066: contig of 3501 bp in length  
32067 32185: gap of unknown length  
32186 36196: contig of 4031 bp in length  
36197 36296: gap of unknown length  
36297 41563: contig of 5267 bp in length  
41564 41663: gap of unknown length  
41664 45475: contig of 3812 bp in length  
45476 45575: gap of unknown length  
45576 48689: contig of 3114 bp in length  
48690 48789: gap of unknown length  
48790 53726: contig of 4937 bp in length  
53727 53826: gap of unknown length  
53827 60447: contig of 6621 bp in length  
60448 60547: gap of unknown length  
60548 67085: contig of 6538 bp in length  
67086 67185: gap of unknown length  
67186 71628: contig of 4443 bp in length  
71629 71728: gap of unknown length  
71729 76314: contig of 4586 bp in length  
76315 76414: gap of unknown length  
76415 82810: contig of 6396 bp in length  
82811 82910: gap of unknown length  
82911 87615: contig of 4705 bp in length  
87616 87715: gap of unknown length  
87716 94295: contig of 6580 bp in length  
94296 94395: gap of unknown length  
94396 99025: contig of 4630 bp in length  
99026 99125: gap of unknown length  
99126 104801: contig of 5676 bp in length  
104802 104901: gap of unknown length  
104902 111798: contig of 6897 bp in length  
111799 111898: gap of unknown length  
111899 121365: contig of 9467 bp in length

\* 121366 121465: gap of unknown length  
\* 121466 129872: contig of 8407 bp in length  
\* 129873 129972: gap of unknown length  
\* 129973 140459: contig of 10487 bp in length  
\* 140460 140559: gap of unknown length  
\* 140560 149383: contig of 8824 bp in length  
\* 149384 149483: gap of unknown length  
\* 149484 159802: contig of 10319 bp in length  
\* 159803 159902: gap of unknown length  
\* 159903 172623: contig of 12721 bp in length.

FEATURES  
Location/Qualifiers

## source

1. .172623  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="rp43-18c12"  
/clone\_lib="RPCI - 43 Male Chimpanzee BAC Library"

## ORIGIN

Query Match 13.0%; Score 40.6; DB 2; Length 172623;  
Best Local Similarity 48.1%; Pred. No. 1.3;  
Matches 115; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 40 GGAAAAAGTTGGTAATACGGAAAAAGAGATTTCATCCAGAAAAAGGTGTGAAAAAGTT 99  
Db 107763 GGCAGACTTCCAGATACAAAAAGGGTAGTGGGAATCTTGCCCATCTGGCATGGAGACGCT 107704  
QY 100 GTGATTGCGTGGAAAAAGTTCCGATCTGACCATCTCTAGATCGTGGAAAAAGTTTCAGTT 159  
Db 107703 CAGCCTAGCATGCGCCCAAGCTGGAATCTTCTGCACAAACAGGTCTGTAGCTTTTTCAT 107644  
QY 160 AGCGCTTACGTACATATGTGGATTGTGAAAAAGAGACGGAGGCATCGGTGAAAAAGA 219  
Db 107643 CATTTTAAAGTGAATTTGAGGAGTGGACACCCAGAGATCATTTTTCAGTCAATG 107584  
QY 220 AGCTTGTCGCTGACGTGACGATAGATAGATACACGTGACGCGTCCACTTTGACGCA 278  
Db 107583 AATTTAAGCAAGATACGGAATAGTTAGATCCAGGTTCTCACGCTGCTATTGACCA 107525

## RESULT 11

AC145064 197004 bp DNA linear HTG 23-APR-2004  
LOCUS Pan troglodytes clone rp43-45h3, WORKING DRAFT SEQUENCE, 11  
DEFINITION unordered pieces.

## ACCESSION

AC145064 GI:46518614

## VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

## SOURCE

Pan troglodytes (Chimpanzee)

## ORGANISM

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

1 (bases 1 to 197004)

Lau,C.C.Y. and Roe,B.A.

Unpublished

2 (bases 1 to 197004)

Lau,C.C.Y. and Roe,B.A.

Direct Submission

Submitted (05-JUN-2003) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

3 (bases 1 to 197004)

Lau,C.C.Y. and Roe,B.A.

Direct Submission

Submitted (23-APR-2004) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

On Apr 23, 2004 this sequence version replaced gi:40018761.

----- Genome Center

Center: Department Of Chemistry And Biochemistry

The University Of Oklahoma

Center code:UOKNOR

-----									
* NOTE: This is a 'working draft' sequence. It currently									
* consists of 11 contigs. The true order of the pieces									
* is not known and their order in this sequence record is									
* arbitrary. Gaps between the contigs are represented as									
* runs of N, but the exact sizes of the gaps are unknown.									
* This record will be updated with the finished sequence.									
* as soon as it is available and the accession number will									
* be preserved.									
		1	3053:	contig of 3053 bp in length					
			3054:	gap of unknown length					
			3153:	gap of unknown length					
			3154:	contig of 3029 bp in length					
			6182:	gap of unknown length					
			6183:	gap of unknown length					
			6283:	contig of 4048 bp in length					
			10330:	contig of 4048 bp in length					
			10331:	gap of unknown length					
			10430:	gap of 7162 bp in length					
			10431:	contig of 7162 bp in length					
			17592:	gap of unknown length					
			17593:	gap of unknown length					
			17692:	contig of 9770 bp in length					
			27462:	gap of unknown length					
			27463:	gap of unknown length					
			27562:	contig of 14115 bp in length					
			41677:	gap of unknown length					
			41678:	gap of unknown length					
			41777:	gap of unknown length					
			57017:	contig of 15240 bp in length					
			57018:	gap of unknown length					
			57117:	gap of unknown length					
			76985:	contig of 19868 bp in length					
			76986:	gap of unknown length					
			77085:	gap of unknown length					
			102682:	contig of 25597 bp in length					
			102683:	contig of unknown length					
			102782:	gap of unknown length					
			135910:	contig of 33128 bp in length					
			135911:	gap of unknown length					
			136010:	gap of unknown length					
			136011:	contig of 60994 bp in length.					
FEATURES									
Location/Qualifiers									
1..197004									
/organism="Pan troglodytes"									
/mol_type="genomic DNA"									
/db_xref="taxon:9598"									
/clone="rp43-45h9"									
/clone_lib="rPCI - 43 Male Chimpanzee BAC Library"									
ORIGIN									
Query Match									
Best Local Similarity 48.1%; Pred. No. 1.3;									
Matches 115; Conservative 0; Mismatches 124; Indels 0; Gaps 0;									
QY	40	GGAAAGTTGGTAATACGGAAGAAAGAGATTTCATCCAGAAAGGTGTGGAAGTT	99						
Db	140399	GGCAGACTTCAGAAATACAAAGGGTAGTGGGAATCTTGCCATCTGGCATGGACGCT	140458						
QY	100	GTGGATTGCGTGAAGAAAGTTTCGATCTGACCATCTCTAGATCGTGGAAGATTCACGTT	159						
Db	140459	CACCTAGCATGCCCCAGCTGGAATCTTCTGCACACAGCTCTGTAGCTTCTTCAT	140518						
QY	160	AGCGCTTACGTCATATGTGGATTGTGGAAAAAGACGGAGGCATCGGTGAAAAAGA	219						
Db	140519	CATTTTAACTGAATTTAGGAGTGTGGAACGCCAGAGATCATTTGTTTCAGCTCAATG	140578						
QY	220	AGCTTGACGCTGACGCTGACATAGATAGATACAGTCGACGCGCTCCACTTGACCA	278						
Db	140579	AATTTAAGCAAGATACGGAATGTAGTTAGATCCAGGTTCTCACGTTCTATTGACCCA	140637						
RESULT 12									
AP004619/c									
LOCUS									
DEFINITION									
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,									
PAC clone:P0583B06.									
AP004619									
ACCESSION									
VERSION									
AP004619.3 GI:38636968									
KEYWORDS									
Oryza sativa (japonica cultivar-group)									
SOURCE									
ORGANISM									
Oryza sativa (japonica cultivar-group)									
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;									
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;									
Ehrhartoideae; Oryzeae; Oryza.									

REFERENCE		
AUTHORS		1 Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE		Oryza sativa nipponbare (GAI) genomic DNA, chromosome 8, PAC clone:P0583B06
JOURNAL		Published Only in Database (2002)
REFERENCE		2 (bases 1 to 134442)
AUTHORS		Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE		Direct Submission
JOURNAL		Submitted (10-JAN-2002) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
		(E-mail: taasaki@ias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT		On Dec 2, 2003 this sequence version replaced gi:32335575. Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/), GlimmerM (http://www.tigr.org/tdb/glimmer/glmr form.html), RiceHMM (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/ep.cgi), gap2 (http://globin.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DBJ. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no.
		A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to RGP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.
		The orientation of the sequence is from SP6 to T7 of the PAC clone. This sequence of P0583B06 clone has an overlap with QJ1134.B10 (DBJ: AP003882) clone at 5' end and with QJ1120.C08 (DBJ: AP003878) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES		Location/Qualifiers
source		1..134442
		/organism="Oryza sativa (japonica cultivar-group)"
		/mol_type="genomic DNA"
		/cultivar="Nipponbare"
		/db_xref="taxon:39947"
		/chromosome="8"
		/clone="P0583B06"
gene		3644..4946
		/gene="P0583B06.1"
misc_feature		3644..4946
		/gene="P0583B06.1"
		/note="probably inactive due to 5' exon missing in CDS probably inactive due to including stop codon(s) in CDS pseudogene, SERK2 protein"
gene		5237..6752
		/gene="P0583B06.2"
misc_feature		5237..6752
		/gene="P0583B06.2"
		/note="probably inactive due to 5' exon missing in CDS pseudogene, SERK1 protein"
gene		complement(14120..17601)
		/gene="P0583B06.3"

misc_feature	complement(14120..17601) /gene="P0583B06.3" /notes="probably inactive due to including frameshift(s) in CDS probably inactive due to including stop codon(s) in CDS pseudogene, RNA-directed DNA polymerase homolog" /pseudo	CDLHYLLIWNDCPLGSLDSWCPVTLSLREFOIEYGSIANVPRKWNMLACLTELDLTLL CSTKQBDIDILGEIIPALLVRLTTHGNGTRIFISSYNAFRCLKYFLHINMGCPNLEE FEGSMQPKLQHLMIKFNHNRWKLNGASDLGRHLSNLTAWEVLICTDYSHGCPNPEE ELMKSTNFAASLIQSARETLPNRPVLRQLOREACIQFEERSVNYEIVNLGEVRDV GKVDQETRVGSKTEPLTYIFEVLMDFIDRFITK" 31852..32823 /gene="P0583B06.7" /join(<31852..31901,32460..>32823) /gene="P0583B06.7" /notes="start and end point are not identified" /join(31852..31901,32460..32823) /gene="P0583B06.7" /notes="predicted by GENSCAN etc." /codon_start=1 /product="hypothetical protein" /protein_id="BAD03232.1" /db_xref="GI:38636972"
gene	complement(18030..19769) /gene="P0583B06.4" 19646..>19769) /gene="P0583B06.4" /notes="start and end point are not identified" complement(join(18030..18299,18388..19352,19433..19546, 19646..19769)) /gene="P0583B06.4" /notes="predicted by GeneMark.hmm etc." /codon_start=1 /product="hypothetical protein" /protein_id="BAD03229.1" /db_xref="GI:38636969"	gene mRNA CDS
gene	complement(19819..21046) /gene="P0583B06.5" 21038..>21046) /gene="P0583B06.5" /notes="start and end point are not identified" complement(join(19819..20120,20296..20499,20584..20686, 21038..21046)) /gene="P0583B06.5" /notes="predicted by GeneMark.hmm etc." /codon_start=1 /product="hypothetical protein" /protein_id="BAD03230.1" /db_xref="GI:38636970"	gene mRNA CDS
gene	complement(21647..29582) /gene="P0583B06.6" join(<21647..22527,27443..>29582) /gene="P0583B06.6" /notes="start and end point are not identified" join(21647..22527,27443..29582) /gene="P0583B06.6" /codon_start=1 /product="putative NBS-LRR disease resistance protein homologue" /protein_id="BAD03231.1" /db_xref="GI:38636971"	Query Match 12.6%; Score 39.2; DB 8; Length 134442; Best Local Similarity 50.0%; Pred. No. 3.2; Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0; QY 29 AATACCATTTGTGAAAAAGTTGGTAATACGAAAAAGAAAGATTCATCATCCAGAAAAAGGT 88 DB 125822 AAAAAAAGTAATAAGATTGTCTTCTAAGGATTTGAAGATAGCTACAAAAAAGG 125763 QY 89 GTGAAAAAGTTGTGGATTGCGTGGAAAAAGTTTCGATCTGACCATCTCTAGATCGTGAAA 148 DB 125762 GTACAAAAAGTTGTGTCGGAGTGGAAAAAATTTGTTGAGTGGACTTTATTGTGAATTA 125703 QY 149 AGTTTCAGTTAGCCCTTACGTACATATCTGATTGTGAAAAAGAAAGCGAGGCATCG 208 DB 125702 AAGGAGAAATTAACAGGGGCATTTTGTGTTGTAATTTGTAATAATAATGGAAG 125643 QY 209 GTGAAAAAGAAAGCTT 224 DB 125642 GATTGAAAAATAGCTT 125627
LOCUS	AP003878	AP003878
DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
ACCESSION	AP003878	BAC clone:OJ1120_C08.
VERSION	AP003878.3	GI:38636796
KEYWORDS		

SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.		
REFERENCE	1		
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.	gene	/gene="OJ1120_C08.2"
TITLE	Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC clone:OJ1120_C08	misc_feature	/note="probably inactive due to including frameshift(s) in CDS
JOURNAL	Published Only in Database (2001)		pseudogene, transposase"
REFERENCE	2 (bases 1 to 150300)		/pseudogene
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.		complement(14189..19746)
TITLE	Direct Submission	gene	/gene="OJ1120_C08.3"
JOURNAL	Submitted (09-JUL-2001) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan	misc_feature	/complement(join(14189..14750,14840..14920,15600..15671,16144..16221,17164..17292,17710..17790,17912..18053,18520..18713,18810..18875,18956..19042,19174..19224),19556..19746))
	IE-mail:tsasaaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)		/note="supported by full-length cDNA(s): AK111670"
COMMENT	On Dec 2, 2003 this sequence version replaced gi:32335570. Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://opal.softberry.com/), GeneMark.hmm (http://www.tigr.org/tcb/glimmer/glmr form.html), RiceHMM (http://www.tigr.org/tcb/glimmer/glmr form.html), SlicerPredictor (http://rgp.dna.affrc.go.jp/RiceHMM/) _SlicerPredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://globin.cse.psu.edu/html/docs/sim4.html), Gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologues of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence. The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of OJ1120_C08 clone has an overlap with P0539806 (DDBJ: AP004619) clone at 5' end and with P0577606 (DDBJ: AP004590) clone at 3' end. The sequence was generated by combining Monsanto and RGP-Japan sequencing data. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.	CDS	/note="probably inactive due to including frameshift(s) in CDS
FEATURES	Location/Qualifiers		
source	1..150300		
	/organism="Oryza sativa (japonica cultivar-group)"		
	/mol_type="genomic DNA"		
	/cultivar="Nipponbare"		
	/db_xref="taxon:39947"		
	/chromosomes="8"		
	/clone="OJ1120_C08"		
gene	5270..8150		
	/gene="OJ1120_C08.1"		
misc_feature	5270..8150		
	/gene="OJ1120_C08.1"		
	/note="probably inactive due to 3' exon missing in CDS		
	probably inactive due to 5' exon missing in CDS		
	pseudogene, MLA6 protein"		
	/pseudogene		
gene	10283..12816		
	/gene="OJ1120_C08.2"		
misc_feature	10283..12816		
	/note="probably inactive due to including frameshift(s) in CDS		
	pseudogene, transposase"		
	/pseudogene		
	complement(14189..19746)		
	/gene="OJ1120_C08.3"		
	/complement(join(14189..14750,14840..14920,15600..15671,16144..16221,17164..17292,17710..17790,17912..18053,18520..18713,18810..18875,18956..19042,19174..19224),19556..19746))		
	/note="supported by full-length cDNA(s): AK111670"		
	/complement(join(14703..14750,14840..14920,15600..15671,16144..16221,17164..17292,17710..17790,17912..18053,18520..18713,18810..18875,18956..19042,19174..19224))		
	/note="contains EST(s): C97094(C51661), AU108860(C51661)		
	contains full-length cDNA(s): AK111670"		
	/codon_start=1		
	/product="putative RAE1 (RNA export 1, S.pombe) homolog"		
	/protein_id="BAD03038.1"		
	/db_xref="GI:38636797"		
	/translation="MATGLAPNMNPNKSFVLPNPGDSVSLSPKSNLLVATSWD		
	NOVRCWEIVGGSSOPKASISHDQPVLCSAWKDDGTTFVSGCDKVKWPLLSGGQAO		
	TVAMHDAVKSIAMIPQNNLLVSSGWDKTLRYWDRQSNVHVQQLPERCYALTYNYP		
	LIMVGTARNLVFNLPQTEFKRIQSLFKYQTRCLAAFPDQQGLVGSIEGRGVGH		
	HWDSQGNKFTKQREGNDIFSVNSLNFHPVHHFTATAGSDAFNFDKDKQRLK		
	AFSCPLIPGCTFNSDGSIFAYAVCYDWSRGAENHPATATSIYLHSPQSEVYKKG		
	PRIATSRK"		
	complement(21178..21432)		
gene	/gene="OJ1120_C08.4"		
misc_feature	complement(21178..21432)		
	/gene="OJ1120_C08.4"		
	/note="hypothetical ORF		
	predicted by GlimmerM		
	this category is not included in IRGSP standard"		
gene	join(21808..21835,21998..22197)		
	/gene="OJ1120_C08.5"		
misc_feature	join(21808..21835,21998..22197)		
	/gene="OJ1120_C08.5"		
	/note="hypothetical ORF		
	predicted by GlimmerM		
	this category is not included in IRGSP standard"		
gene	complement(join(23249..24465,24547..24604,24703..24780,24880..24939,25067..25180,25284..25730))		
	/gene="OJ1120_C08.7"		
misc_feature	complement(join(23249..24465,24547..24604,24703..24780,24880..24939,25067..25180,25284..25730))		
	/gene="OJ1120_C08.7"		
	/note="supported by full-length cDNA(s): AK107028"		
gene	23501..24827		
	/gene="OJ1120_C08.6"		
misc_feature	<23501..>24827		
	/gene="OJ1120_C08.6"		
	/note="supported by full-length cDNA(s): AK107399"		
	23501..24827		
	/gene="OJ1120_C08.6"		
	/note="contains full-length cDNA(s): AK107399		
	non-coding transcript		
	probably inactive due to including stop codon(s) in CDS"		
CDS	complement(join(23819..24465,24547..24604,24703..24780,24880..24939,25067..25180,25284..25496))		
	/gene="OJ1120_C08.7"		
	/note="contains full-length cDNA(s): AK107028, AK109294"		
	/codon_start=1		
	/product="putative transcription factor STGAL"		
	/protein_id="BAD03039.1"		
	/db_xref="GI:38636798"		
	/translation="MTSAAPAAQAFAPAPAMGIYDRRHHPHPLAAGVGDHPFIRPDT		
	TASTNSAAAAAMVVPPLTPKFSQALPLQHGDDQDAAALQESPRHADSFEQE		
	ASKPRDKIQRRLAQNREARKYRIQNLETSRMKLAHLEQETIRARQOSYAIN		
	RSSNPATLPAPIDSGVTFVEYQAQWVEEQGQTAEELRASLQAAAEGLRAVVEAAL		

```
AHYDRFLPAKREARDVFFVMSGVWRTGAERFLWIAGFRPSEVIRVLAQLEPMTWTE
ROADYVQGLQKARHSDALSGQMDLKTQLADSLIAEAVVWVSTCDSPPPPPPEE
EPSSAAGDGGCYMAQMGKSLNLVAFVDHVRHRRSPPTSHLHVRRAELG"
gene      complement (join(29680..29827,31003..31098,32215..32219,
33217..33312))
/genes="OJ1120 C08.8"
misc_feature complement (join(29680..29827,31003..31098,32215..32219,
33217..33312))
/genes="OJ1120 C08.8"
/notes="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
gene      join(37583..37663,37864..38022)
/genes="OJ1120 C08.9"
misc_feature join(37583..37663,37864..38022)
/genes="OJ1120 C08.9"
/notes="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
gene      complement (38722..38910)
/genes="OJ1120 C08.10"
mRNA      complement (<38722..>38910)
/genes="OJ1120 C08.10"
/notes="start and end point are not identified"
CDS        complement (38722..38910)
/genes="OJ1120 C08.10"
/notes="predicted by FGENESH etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BA003040.1"
/db_xref="GI:3836799"
/translations="MYVEAAVSSLKQQQQQQQRRGHKSAARSHQVSIKKEQQQQIG
CRTLWAPRIYPTASGKK"
gene      complement (38954..39154)
/genes="OJ1120 C08.11"
misc_feature complement (38954..39154)
/genes="OJ1120 C08.11"
/notes="hypothetical ORF
predicted by GlimmerM
this category is not included in IRGSP standard"
gene      complement (join(41298..41610,42041..43471,44688..45078,
45161..45373,45480..45520))
/genes="OJ1120 C08.12"
mRNA      complement (join(41298..41610,42041..43471,44688..45078,
45161..45373,45480..45520))
/genes="OJ1120 C08.12"
/notes="supported by full-length cDNA(s): AK105561"
CDS        complement (join(42104..43471,44688..45078,45161..45351))
Query Match      12.6%; Score 39.2; DB 8; Length 150300;
Best Local Similarity 50.0%; Pred. No. 3.2;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 29 AATACCATGTGGAAGTGTGTAATACGGAAGAAAGAGATTCATCCAGAAAGGT 88
|||
Db 83780 AAAAAAAGTAAATAAGATTGCTATTGCTAAGGATTGCAATAGTACAAAAAGG 83721
|||

QY 89 GTGGAAGTGTGGATTGGTGGGAAGATTGCATTCGATTCGATTCGTAGCTGGA 148
|||
Db 83720 GTAACAAAGTGTGTCGGAGTGGAAACAAATTTGTTGAGTGGACATTATTTGAATTA 83661
|||

QY 149 AAGTTCACGTAGCGCTTACGTACATATGTTGATGTTGGAAGAAAGAGCGAGGCATCG 208
|||
Db 83660 AAGAGNAATTACAGGGGATTTTCTTGCATTGTGAAGAAAGTAATAATGAAG 83601
|||

QY 209 GTGGAAGAAAGACTT 224
|||
Db 83600 GATTGGAAGAAAGCTT 83585
|||

RESULT 14
AC112060/c
LOCUS          AC112060          182932 bp      DNA      linear      HTG 15-NOV-2002
```

```
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
```

```
Rattus norvegicus clone CH230-180E21, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC112060
HTG1; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 182932)
Muzny, D., Marie, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Aisbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, R.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhenwa, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naik, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Naokelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reich, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 182932)
Worley, K. C.
Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 182932)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
```

COMMENT

Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:22856848.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information  
Center project name: GQFS  
Center clone name: CH230-180E21  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 171712 bases at least Q40  
Consensus quality: 174959 bases at least Q30  
Consensus quality: 177170 bases at least Q20  
Estimated insert size: 183207; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 179536: contig of 179536 bp in length  
\* 179537 179636: gap of unknown length  
\* 179637 181302: contig of 1666 bp in length  
\* 181303 181402: gap of unknown length  
\* 181403 182932: contig of 1530 bp in length.

FEATURES  
source  
1. 182932  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-180E21"  
misc\_feature  
1. 1190  
/notes="wgs contig"  
misc\_feature  
177608..179536  
/note="wgs\_contig"

ORIGIN

Query Match 12.6%; Score 39.2; DB 2; Length 182932;  
Best Local Similarity 49.5%; Pred. No. 3.2;  
Matches 101; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 17 AAGGGCTCGTATACCATTTGGAAAAAGTTGGTAATACGGAAGAAAGAGATTTCATCA 76  
Db 180078 AAAGACATGAAGAACTCCCTTAGGGAACACAGGAAACATTAATAACAGTAGAAGCC 180019  
Qy 77 TCCAGAAAAGCTGTGAAAAGTTGTGGATTGGCGTGAAGAAAGTTCCGATCTGACCATCTCT 136  
Db 180018 TACAGAGAGAAATCGCAAAATCCCTGAAGAAATCGCAAAATGCTGAAGAAATCCA 179959  
Qy 137 AGATCGTGGAAAAGTTACGTTAGCGCTTACGTACATATGTGGATTTGGAAAAAGAG 196  
Db 179958 GGAAGAACACACAGTTGAAGGAACATAAAATGGAATAGAACCAATCAAGAAAGAACAC 179899

Qy 197 ACGGAGGCGATCGTGGAAAAAGAA 220  
Db 179898 ATGGAACAACCCCTGGATATAGAA 179875

RESULT 15  
AC108612  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-293B15, WORKING DRAFT SEQUENCE, 2  
unordered pieces.  
AC108612  
AC108612.5 GI:25139237  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
ORGANISM  
Rattus.  
1 (bases 1 to 185618)  
Muzny, D. Marie., Metzker, M. Lee., Abranzon, S., Adams, C., Alder, J.,  
Allen, C., Allien, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. O., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escoto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Greggorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunarathne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuhewa, L., Louissegh, H., Lozado, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
Nwaokeme, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.  
Direct Submission  
Unpublished  
2 (bases 1 to 185618)  
Worley, K. C.  
Direct Submission

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE



## JOURNAL

Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 185618)  
 Rat Genome Sequencing Consortium.  
 Direct Submission

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Nov 20, 2002 this sequence version replaced gi:22856935.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GP00

Center clone name: CH230-293B15

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 178343 bases at least Q40

Consensus quality: 179378 bases at least Q30

Consensus quality: 180026 bases at least Q20

Estimated insert size: 180394; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 184363: contig of 184363 bp in length

\* 184364 184463: gap of unknown length

\* 184464 185618: contig of 1155 bp in length.

FEATURES

## source

Location/Qualifiers

1. .185618

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-293B15"

1. .1155

/note="wgs contig"

misc\_feature

183070..184363

/note="wgs\_contig"

## ORIGIN

Query Match 12.6%; Score 39.2; DB 2; Length 185618;  
 Best Local Similarity 49.5%; Pred. No. 3.3;  
 Matches 101; Conservative 0; Mismatches 103; Indels 0; Gaps 0;  
 QY 17 AAGGGCTCGGTAAATACATTGTGGAAAAAGTTGGTAATACGGAAGAAAGAAATTCATCA 76  
 Db 138350 AAGACATGATGAACTCCCTTAGGGAACACACGAAACACATTATTAACAGTAGAGCC 138409  
 QY 77 TCCAGAAAAGGTGTGAAAAAGTTGTGGATTGCGTGGAAAAAGTTGCGATCTGACCATCTCT 136

Db 138410 TACAGAGAGGAATCGCAAAAAATCCCTGAAAGAAATCGCAAAAAATGCTGAAAGAATTCCA 138469  
 QY 137 AGATCGTGGAAAAAGTTTCAGCTTAGCGCTTACGTCACATATGTGGATTGTGAAAAAGAAG 196  
 Db 138470 GGAACACACACAGCTTGAAGGAACCTAATAATGGAATAGAACCAATCAAGAAAGAACAC 138529  
 QY 197 ACGGAGGCGATCGGTGAAAAAGAA 220  
 Db 138530 ATGGAACACACCCCTGGATATAGAA 138553

Search completed: July 3, 2005, 03:19:05  
 Job time : 1737.41 secs

Page blank (usp:0)



```
PF 25-FEB-1999; 99EP-00301419.
XX
PR 09-NOV-1998; 98IN-DE003322.
XX
PA (COUL ) CSIR COUNCIL SCI IND RES.
PI Tuli R, Sawant SV, Singh PK, Gupta SK;
XX WPI; 2000-341712/30.
XX
XX New chemically synthesized artificial promoter, useful high level
PT expression of transgenes in different organisms.
XX
XX Claim 13; Page 11; 40pp; English.
XX
XX Chemically synthesized artificial promoters are new and comprise a DNA
CC sequence designed for a targeted level and pattern of gene expression by
CC strategically putting together several signature sequences identified by
CC sequence alignment and statistical analysis of a large database
CC constructed for this purpose. A method for chemically synthesizing an
CC artificial promoter for expressing genes at a desired level in different
CC organisms is also claimed. The high level expression in a plant using
CC such an artificial promoter (e.g. AAA28449) can be measured comprising
CC polyethylene glycol (PEG) mediated transformation of plant protoplasts as
CC well as biolistic mediated transformation of plant tissues including
CC root, stem, intact leaf tissue followed by transient GUS assay to compare
CC with a natural CamV 35S promoter showing the desired level of activity.
CC The promoter is useful for high level expression of transgenes in
CC different organisms and for testing high level gene expression in plants
CC (claimed). The promoter is biologically active and is efficient and can
CC be synthesized to express in even the most complex organisms
XX
XX Sequence 452 BP; 157 A; 88 C; 95 G; 112 T; 0 U; 0 Other;
XX
Query Match 99.5%; Score 310.4; DB 3; Length 452;
Best Local Similarity 99.7%; Pred. No. 2e-87;
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTCGACCATCATTTGAAAGGCCCTCGGTATACATCATTTGTGGAAGTTGGTAATACGGA 60
DB 1 GTCGACCATCATTTGAAAGGCCCTCGGTATACATCATTTGTGGAAGTTGGTAATACGGA 60
QY 61 AAAAGAGATTTCATCATCCAGAAAGGTGTGGAAGTTCTCGATTGCGTGGAAAAGTT 120
DB 61 AAAAGAGATTTCATCATCCAGAAAGGTGTGGAAGTTCTCGATTGCGTGGAAAAGTT 120
QY 121 CGATCTGACCATCTCTAGATCGTGGAAAAGTTTCAGCTTAGCGCTTACATATGTGG 180
DB 121 CGATCTGACCATCTCTAGATCGTGGAAAAGTTTCAGCTTAGCGCTTACATATGTGG 180
QY 181 ATTGTGGAAGAAAGACGAGCGCATCGGTGGAAAAGCTTGTACGCTGTACGCTGA 240
DB 181 ATTGTGGAAGAAAGACGAGCGCATCGGTGGAAAAGCTTGTACGCTGTACGCTGA 240
QY 241 CGATAGATAGATACGTCACGCGTCCACTTGACCAATTCAGCGCAATGACGCCA 300
DB 241 CGATAGATAGATACGTCACGCGTCCACTTGACCGCAATTCAGCGCAATGACGCCA 300
QY 301 CTTGACGCTACT 312
DB 301 CTTGACGCTACT 312
```

## RESULT 2

AAA28438

ID AAA28438 standard; DNA; 46 BP.

AC AAA28438;

XX 29-AUG-2000 (first entry)

XX Synthetic promoter conserved domain I.

```
KW Artificial promoter; strong; weak; transgene expression; plant; ss.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FH misc_signal 1..21
FT /tag= a
FT /label= U_box
FT /note= "GTACGC type element"
FT repeat_region 22..33
FT /tag= b
FT /rpt_type= TANDEM
FT repeat_unit 22..25
FT /tag= c
FT /note= "GATA-type cis-acting element"
FT GC_signal 34..46
FT /tag= d
FT /note= "Resembles GC-box, may play role in kinetics of
FT opening of the transcription bubble and keeping the
FT minimal promoter in a most active form to enhance
FT transcription reinitiation"
XX
XX BP1002869-A1.
XX
XX 24-MAY-2000.
XX
XX 25-FEB-1999; 99EP-00301419.
XX
XX 09-NOV-1998; 98IN-DE003322.
XX
XX (COUL ) CSIR COUNCIL SCI IND RES.
XX
XX Tuli R, Sawant SV, Singh PK, Gupta SK;
XX WPI; 2000-341712/30.
XX
XX New chemically synthesized artificial promoter, useful high level
PT expression of transgenes in different organisms.
XX
XX Claim 5; Page 16; 40pp; English.
XX
XX A chemically synthesized promoter can comprise a conserved domain I as
CC shown here for high level expression of genes. Chemically synthesized
CC artificial promoters are new and comprise a DNA sequence designed for a
CC targeted level and pattern of gene expression by strategically putting
CC together several signature sequences identified by sequence alignment and
CC statistical analysis of a large database constructed for this purpose. A
CC method for chemically synthesizing an artificial promoter for expressing
CC genes at a desired level in different organisms is also claimed. The high
CC level expression in a plant using such an artificial promoter (e.g.
CC AAA28449) can be measured comprising polyethylene glycol (PEG) mediated
CC transformation of plant protoplasts as well as biolistic mediated
CC transformation of plant tissues including root, stem, intact leaf tissue
CC followed by transient GUS assay to compare with a natural CamV 35S
CC promoter showing the desired level of activity. The promoter is useful
CC for high level expression of transgenes in different organisms and for
CC testing high level gene expression in plants (claimed). The promoter is
CC biologically active and is efficient and can be synthesized to express in
CC even the most complex organisms
XX
XX Sequence 46 BP; 11 A; 11 C; 13 G; 11 T; 0 U; 0 Other;
XX
Query Match 14.7%; Score 46; DB 3; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 221 GCTTGTACGCTGTACGCTGACGATAGATAGATACACGTCACGCGT 266
DB 1 GCTTGTACGCTGTACGCTGACGATAGATAGATACACGTCACGCGT 46
```

RESULT 3

AAA28437

AA28437 standard; DNA; 46 BP.  
AA28437;  
29-AUG-2000 (first entry)  
Synthetic promoter minimal domain (b) for high level gene expression.  
Artificial promoter; strong; weak; transgene expression; plant; ss.  
Synthetic.  
Key Location/Qualifiers  
CAAT\_signal 1..5  
/\*tag= a  
/note= "conserved"  
misc\_signal 6..10  
/\*tag= b  
/note= "conserved TGACG box"  
CAAT\_signal 11..16  
/\*tag= c  
/note= "conserved"  
misc\_signal 17..21  
/\*tag= d  
/note= "conserved TGACG box"  
CAAT\_signal 22..27  
/\*tag= e  
/note= "conserved"  
misc\_signal 27..31  
/\*tag= f  
/note= "conserved TGACG box"  
CAAT\_signal 32..36  
/\*tag= g  
/note= "conserved"  
misc\_signal 37..41  
/\*tag= h  
/note= "conserved TGACG box"  
CAAT\_signal 42..46  
/\*tag= i  
/note= "conserved"  
EPI002869-A1.  
24-MAY-2000.  
25-FEB-1999; 99EP-00301419.  
09-NOV-1998; 98IN-DE003322.  
(COUL ) CSIR COUNCIL SCI IND RES.  
Tuli R, Sawant SV, Singh PK, Gupta SK;  
WPI; 2000-341712/30.  
New chemically synthesized artificial promoter, useful high level  
expression of transgenes in different organisms.  
Claim 4; Page 15; 40pp; English.  
A chemically synthesized promoter can comprise a minimal domain (b) as  
shown here for high level expression of genes. This domain has a number  
of conserved sequences interspersed with a TGACG box. Chemically  
synthesized artificial promoters are new and comprise a DNA sequence  
designed for a targeted level and pattern of gene expression by  
strategically putting together several signature sequences identified by  
sequence alignment and statistical analysis of a large database  
constructed for this purpose. A method for chemically synthesizing an  
artificial promoter for expressing genes at a desired level in different  
organisms is also claimed. The high level expression in a plant using  
such an artificial promoter (e.g. AA28449) can be measured comprising  
polyethylene glycol (PEG) mediated transformation of plant protoplasts as  
well as biolistic mediated transformation of plant tissues including

CC root, stem, intact leaf tissue followed by transient GUS assay to compare  
CC with a natural CamV 35S promoter showing the desired level of activity.  
CC The promoter is useful for high level expression of transgenes in  
CC different organisms and for testing high level gene expression in plants  
CC (claimed). The promoter is biologically active and is efficient and can  
CC be synthesized to express in even the most complex organisms  
XX  
SQ Sequence 46 BP; 13 A; 16 C; 8 G; 9 T; 0 U; 0 Other;  
Query Match 14.7%; Score 46; DB 3; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.00017;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 267 CCACCTTGACGCAACAATTGACGCAATGACGCCACTTGACGCTACT 312  
DB 1 CCACTTGACGCAACAATTGACGCAATGACGCCACTTGACGCTACT 46  
RESULT 4  
ACN45090\_1  
Continuation (2 of 4) of ACN45090 from base 100001 (Human genomic sequence hCG22125.)  
WP Sequence split into 4 fragments LOCUS ACN45090 Accession ACN45090  
WP Fragment Name Begin End  
WP ACN45090\_0 1 110000  
WP ACN45090\_1 100001 210000  
WP ACN45090\_2 200001 310000  
WP ACN45090\_3 300001 350764  
Query Match 11.8%; Score 36.8; DB 11; Length 110000;  
Best Local Similarity 56.7%; Pred. No. 2.7;  
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 70 TTCATCATCAGAAAGGTTGGAAAAGTTGTGGATTGCGTGGAAAAGTTTCGATCTGAC 129  
DB 76489 TTCTTTTAAAAAATGATGCAAAAGCTGTGCTGGCGGGGCTATTTTACATCGAA 76548  
QY 130 CATCTCTAGATCGTGGAAAAGTTACGTTAGCGCTTACGTACATATGTGGATTGTGGAA 189  
DB 76549 ATGCTTAAGAGATTGTAATAAGGTTGGTGAAGAACTGAGGTACATTTAATGAATATGAA 76608  
RESULT 5  
ABL02730  
ID ABL02730 standard; cDNA; 7450 BP.  
XX AC ABL02730;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2672.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical; gene; ss.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX PR 11-JUL-2000; 2000US-00614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX DR P-PSDB; ABB58627.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 2672; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 7450 BP; 2044 A; 1752 C; 1603 G; 2051 T; 0 U; 0 Other;  
Query Match 11.0%; Score 34.4; DB 4; Length 7450;  
Best Local Similarity 57.4%; Pred. No. 5.5;  
Matches 62; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 11 ATTGAAAGGGCTCGGTATACCAATTGTGGAAGAGTTGGTAATACGGAAGAAAGAT 70  
DB 7068 ATTCTTTGGCCCTGAGTAATTTCAATTAGAAATAAGCGGCANTTAGGGACATGTGAA 7127  
QY 71 TCATCATCCAGAAAGGTGTGGAAGAGTTGTGGATTGCGTGGAAAAAG 118  
DB 7128 AAAAGAGCGGAAACGGTTTTCGAAACATGTTTTGGCCCAAGTTAAAG 7175  
RESULT 6  
ADB69083/c  
ID ADB69083 standard; DNA; 4074 BP.  
XX  
AC ADB69083;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE C. neoformans genomic DNA sequence SEQ ID NO:210.  
XX  
KW ds; gene; fungicide; gene therapy; infection.  
XX  
OS Cryptococcus neoformans.  
XX  
PN WO2003052076-A2.  
XX  
PD 26-JUN-2003.  
XX  
PF 17-DEC-2002; 2002WO-US040225.  
XX  
PR 17-DEC-2001; 2001US-0341261P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Zamudio C, Eroshkin AM;  
XX  
XX WPI; 2003-533017/50.  
XX  
XX P-PSDB; ADB70166.  
XX  
XX New nucleic acid, useful for preparing a composition for treating an  
PT infection caused by Cryptococcus neoformans.  
XX  
XX Claim 3; SEQ ID NO 210; 136pp; English.  
XX  
XX The invention relates to a novel purified or isolated Cryptococcus  
CC neoformans nucleic acid molecule comprising a sequence encoding a  
CC polypeptide comprising a sequence not given in the specification. A  
CC polynucleotide of the invention has fungicide activity, and may have a  
CC use in gene therapy. The nucleic acid is useful for preparing a  
CC composition for treating an infection caused by Cryptococcus neoformans.  
CC The present sequence represents a C. neoformans sequence of the  
CC invention. Note: The sequence data for this patent is not represented in

CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 4074 BP; 1066 A; 930 C; 1005 G; 1070 T; 0 U; 3 Other;  
Query Match 10.6%; Score 33; DB 10; Length 4074;  
Best Local Similarity 52.6%; Pred. No. 12;  
Matches 72; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
QY 16 AAAGGGCTCGGTATACCAATTGTGGAAGAGTTGGTAATACGGAAGAAAGATTCATC 75  
DB 3865 AAATGGTCCCATATATATATTGGTCAAGGACGACATATTCATCAAAAGCACATTTT 3806  
QY 76 ATCCAGAAAGGTGTGGAAGAGTTGTGGATTGCGTGGAAAAAGTTGCATCTGACCATCTC 135  
DB 3805 TGTAAATAAAGGCATGAGAATTTTGAGGCTTTTGTAAACAAAGAGTTGAATTGAACATATG 3746  
QY 136 TAGATCGTGGAAAAAGT 152  
DB 3745 GGGATGGTGATACAAAT 3729  
RESULT 7  
ABK76259  
ID ABK76259 standard; DNA; 468 BP.  
XX  
AC ABK76259;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Bacillus licheniformis genomic sequence tag (GST) #3550.  
XX  
KW Differential gene expression; genomic sequenced tag; GST;  
KW altered culture condition; environmental stress;  
KW physiological provocation; ds.  
XX  
OS Bacillus licheniformis.  
XX  
PN WO200229113-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-US031437.  
XX  
PR 06-OCT-2000; 2000US-00680598.  
XX  
PR 27-MAR-2001; 2001US-0279526P.  
XX  
XX (NOVO ) NOVOZYMES BIOTECH INC.  
XX  
XX (NOVO ) NOVOZYMES AS.  
XX  
XX Berka R, Clausen IG;  
XX  
XX WPI; 2002-416684/44.  
XX  
XX Monitoring differential expression of several genes in first Bacillus  
PT cell relative to expression of same genes in one or more second Bacillus  
PT cells, by using substrate containing Bacillus genomic sequenced tag  
PT array.  
XX  
XX Claim 4; SEQ ID NO 3550; 200pp; English.  
XX  
XX The invention describes a method of monitoring differential expression of  
CC genes in a first Bacillus cell relative to expression of the genes in  
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
CC isolated from Bacillus cells to a substrate containing array of Bacillus  
CC genomic sequenced tags (GST), examining the array, and determining  
CC relative gene expression by an observed hybridisation reporter signal of  
CC a spot in the array. The method is useful for measuring the expression of  
CC genes in a first Bacillus cell relative to expression of the same genes  
CC in one or more second Bacillus cells. The method is useful for monitoring  
CC global expression of several genes from a Bacillus cell, discovering new  
CC genes, identifying possible functions of unknown open reading frames and  
CC monitoring gene copy number variation and stability. Monitoring changes

CC in expression of genes may be used to provide a representation of the way  
CC in which *Bacillus* cells adapt to changes in culture conditions,  
CC environmental stress or other physiological provocation. Extensive follow  
CC -up characterisation is unnecessary, when one spot on an array equals one  
CC gene or one open reading frame, since sequence information is available.  
CC This sequence represents a genomic sequence tag (GST) used in the method  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 468 BP; 124 A; 112 C; 114 G; 118 T; 0 U; 0 Other;

Query Match 10.4%; Score 32.6; DB 6; Length 468;

Best Local Similarity 55.9%; Pred. No. 7;

Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 21 GCCTCGGTATACATTCGTAAGAGTTGGTATCGGTAAGAGAGATTCATCATCCA 80  
DB 191 GTCTCGACAGTCCGGCTTTACACGATTCAAAATCGGGAAGGCGATTTTATATCC 250  
QY 81 GAAAAGGTGTGGAAGTTCTGGATTGCGTGGAAAAAGTTCTGACCA 131  
DB 251 GGAAGGGCTTGAACTTTAGACAGTCCATGGAAAAATGGGAGCTGACAA 301

## RESULT 8

ABQ69245\_01  
Continuation (2 of 31) of ABQ69245 from base 100001 (*Listeria innocua* DNA sequence #684).

WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245

WP Fragment Name Begin End

WP ABQ69245\_00 1 110000

WP ABQ69245\_01 100001 210000

WP ABQ69245\_02 200001 310000

WP ABQ69245\_03 300001 410000

WP ABQ69245\_04 400001 510000

WP ABQ69245\_05 500001 610000

WP ABQ69245\_06 600001 710000

WP ABQ69245\_07 700001 810000

WP ABQ69245\_08 800001 910000

WP ABQ69245\_09 900001 1010000

WP ABQ69245\_10 1000001 1110000

WP ABQ69245\_11 1100001 1210000

WP ABQ69245\_12 1200001 1310000

WP ABQ69245\_13 1300001 1410000

WP ABQ69245\_14 1400001 1510000

WP ABQ69245\_15 1500001 1610000

WP ABQ69245\_16 1600001 1710000

WP ABQ69245\_17 1700001 1810000

WP ABQ69245\_18 1800001 1910000

WP ABQ69245\_19 1900001 2010000

WP ABQ69245\_20 2000001 2110000

WP ABQ69245\_21 2100001 2210000

WP ABQ69245\_22 2200001 2310000

WP ABQ69245\_23 2300001 2410000

WP ABQ69245\_24 2400001 2510000

WP ABQ69245\_25 2500001 2610000

WP ABQ69245\_26 2600001 2710000

WP ABQ69245\_27 2700001 2810000

WP ABQ69245\_28 2800001 2910000

WP ABQ69245\_29 2900001 3010000

WP ABQ69245\_30 3000001 3111208

## Query Match

Best Local Similarity 10.4%; Score 32.6; DB 6; Length 110000;

Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 62 AAGAAGATTATCATCCAGAAAAGTTGTGAAAAGTTGTGGATTTCGTTGGAAGAGTTTC 121

DB 46002 AATAGAGTTTATAATCAAAATCGGGTGAAGAGATGAATAAGAAGATGAAGAGCTT 46061

QY 122 GATCTGACCATCTCTAGATCGTGGAAAAAGTTACGTTAGCGTTACGTACATATGGA 181

DB 46062 TTTAGTATTACTCTCTGTTGTTGTAGTATTATTCGTTAGTGTGTTGATCAATTGTCGA 46121

QY 182 TTGTGGAAGAAAGACGAGGCG 204  
DB 46122 GAGTCAAAACGGGCGAAAGCGGC 46144

## RESULT 9

ABQ67197\_00  
WP Sequence split into 12 fragments LOCUS ABQ67197 Accession Abq67197

WP Fragment Name Begin End

WP ABQ67197\_00 1 110000

WP ABQ67197\_01 100001 210000

WP ABQ67197\_02 200001 310000

WP ABQ67197\_03 300001 410000

WP ABQ67197\_04 400001 510000

WP ABQ67197\_05 500001 610000

WP ABQ67197\_06 600001 710000

WP ABQ67197\_07 700001 810000

WP ABQ67197\_08 800001 910000

WP ABQ67197\_09 900001 1010000

WP ABQ67197\_10 1000001 1110000

WP ABQ67197\_11 1100001 1163020

ID ABQ67197 standard; DNA; 1163020 BP.

XX AC

XX ABQ67197;

XX AC

XX 29-AUG-2002 (first entry)

XX DE *Listeria innocua* contig DNA sequence #10.

XX DE *Listeria innocua* contig DNA sequence #10.

XX KW Antibacterial; *Listeria*; food contamination; mutational analysis;

XX KW Infection; ds.

XX OS *Listeria innocua*.

XX PN WO200228891-A2.

XX PD 11-APR-2002.

XX XX 04-OCT-2001; 2001WO-FR003061.

XX XX 04-OCT-2000; 2000FR-00012697.

XX PA (INSP ) INST PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Kunst F, Glaser P;

XX XX WPI; 2002-332479/37.

XX DR New genomic sequences from *Listeria* species, useful for detection,

XX PT treatment and prevention of infection, also related polypeptides,

XX PT antibodies and modulators.

XX XX Claim 5; SEQ ID NO 10; 180bp; French.

XX CC The present invention relates to nucleic acid sequences (ABQ67189-

XX CC ABQ71212) from *Listeria* sp. The sequences are useful as probes and

XX CC primers for identification and/or detection of *Listeria* (e.g. as

XX CC contaminants in foods, or mutational analysis) and for analysis of gene

XX CC expression. Proteins encoded by the nucleic acid sequences can be used to

XX CC screen for compounds that modulate gene expression, replication and

XX CC pathogenicity of *Listeria* (potential therapeutic agents), also for

XX CC treating infections by *Listeria*, and are useful as immunogens in anti-

XX CC *Listeria* vaccines. Note: The sequence data for this patent did not form

XX CC part of the printed specification, but was obtained in electronic format

XX CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1163020 BP; 388339A; 197589C; 235612G; 341474T; 0U; 60Other;

## Query Match

Best Local Similarity 10.4%; Score 32.6; DB 6; Length 110000;

Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;





XX This sequence encodes a Streptococcus pneumoniae protein that (based on  
 CC homology with a Bacillus subtilis protein) is a BSELABCD NCBI gi:  
 CC 895746, and represents a DNA sequence of the invention. The DNA sequences  
 CC were isolated from Streptococcus pneumoniae strain 0100993 (NCIMB 40794).  
 CC The Streptococcus pneumoniae proteins of the invention can be used to  
 CC identify compounds which interact with and inhibit or activate the  
 CC activity of the proteins. Antagonists can be used to treat diseases  
 CC caused by S. pneumoniae proteins, through genetic immunisation. They can  
 CC also be used to induce an immunological response in a mammal by  
 CC inoculation with the S. pneumoniae proteins or delivery of the encoding  
 CC nucleic acids in a vector adequate to produce antibody and/or T cell  
 CC immune responses to protect the animal from disease. The proteins can  
 CC also be used to identify antimicrobial compounds which are capable of  
 CC inhibiting their bioactivity. In particular the proteins of the invention  
 CC can be used to prevent adhesion of bacteria to mammalian extracellular  
 CC matrix proteins on in-dwelling devices or in wounds, to block protein-of  
 CC mediated mammalian cell invasion, and to block the normal progression of  
 CC pathogenesis in infections initiated other than by the implantation of in  
 CC -dwelling devices or other surgical techniques

XX Sequence 1096 BP; 338 A; 224 C; 138 G; 396 T; 0 U; 0 Other;

Query Match 10.4%; Score 32.4; DB 2; Length 1096;

Best Local Similarity 51.4%; Pred. No. 11;

Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 62 AAGAGAGATTCATCCAGAAAAGGTGTGGAAGTTTGGATTGCGTGGAAAAGTTTC 121

DB 699 AGACAAGGTAAATTTACTTTGAAAAGATGGTTGAGAGTTTATCTGTGGTGAATCAATGA 640

QY 122 GATCTGACCATCTCTAGATCGTGGAAAAGTTTCAACGTTAGCGTTACGTACATATGTGGA 181

DB 639 GTTTGAACAATCTTTACTTTATGGAATTAACAGCGGAGAGATTAGTTTCAGTTGTTT 580

QY 182 TTGTGAAAAGAACGCGGAGGCATC 207

DB 579 TTCTGAAAAAATTCGAGTCCGCATC 554

#### RESULT 12

AAZ96371/C

ID AAZ96371 standard; DNA; 1727 BP.

XX AAZ96371;

DT 10-APR-2000 (first entry)

XX S. pneumoniae derived DNA from ORF #199.

XX Treatment; prevention; disease; diagnosis; gene therapy; screening;  
 KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; sb.

XX Streptococcus pneumoniae.

XX WO9806734-A1.

XX 19-FEB-1998.

XX 15-AUG-1997; 97WO-US014436.

XX 16-AUG-1996; 96US-0024022P.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
 PI Stodola RK;

XX WPI; 1998-159452/14.

XX P-PSDB; AAY86040.

XX Streptococcus pneumoniae proteins and related DNA - useful for screening  
 PT compounds for antibacterial activity.

XX PS

Claim 4; Page 230-231; 640pp; English.

CC This invention describes novel isolated Streptococcus pneumoniae  
 CC polynucleotides (see DNA296173-296494) and their encoded proteins (see  
 CC AAY85792-Y86182). The DNA, vectors and host cells described in the method  
 CC of the invention are useful for the recombinant expression of the  
 CC polypeptides. The polypeptides are useful for treatment or prevention of  
 CC disease, or diagnosis of disease related to expression or activity of  
 CC such a polypeptide. They can also be used to screen for compounds which  
 CC interact with and inhibit or activate such a polypeptide. The  
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful for  
 CC inducing an immunological response in a mammal. The antagonists are  
 CC useful to inhibit such bacterial polypeptides. The polypeptides are  
 CC particularly useful to identify antimicrobial compounds and antibiotics.  
 CC They are also useful to determine their role in pathogenesis of  
 CC infection, dysfunction and disease

XX Sequence 1727 BP; 546 A; 332 C; 221 G; 628 T; 0 U; 0 Other;

Query Match 10.4%; Score 32.4; DB 2; Length 1727;

Best Local Similarity 51.4%; Pred. No. 13;

Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 62 AAGAGAGATTCATCCAGAAAAGGTGTGGAAGTTTGGATTGCGTGGAAAAGTTTC 121

DB 696 AGACAAGGTAAATTTACTTTGAAAAGATGGTTGAGAGTTTATCTGTGGTGAATCAATGA 637

QY 122 GATCTGACCATCTCTAGATCGTGGAAAAGTTTCAACGTTAGCGTTACGTACATATGTGGA 181

DB 636 GTTGAACAATCTTTACTTTATGGAATTAACAGCGGAGAGATTAGTTTCAGTTGTTT 577

QY 182 TTGTGAAAAGAACGCGGAGGCATC 207

DB 576 TTCTGAAAAAATTCGAGTCCGCATC 551

#### RESULT 13

AAI66078

ID AAI66078 standard; DNA; 2992 BP.

XX AAI66078;

XX 14-JAN-2002 (first entry)

XX Spinacia oleracea ClpC protease gene (GenBank: AF043539).

XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;  
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;  
 KW Arabidopsis thaliana; ds.

XX Spinacia oleracea.

XX WO200170929-A2.

XX 27-SEP-2001.

XX 20-MAR-2001; 2001WO-US008836.

XX 20-MAR-2000; 2000US-0190769P.

XX 18-APR-2000; 2000US-0198116P.

XX (ARCH-) ARCH DEV CORP.

XX Lindquist S, Queitsch C, Vierling E;

XX WPI; 2001-639123/73.

XX Transgenic plants with improved heat stress tolerance, useful for  
 PT producing animal feed, oil and synthetic products.

XX Claim 4; Page; 91pp; English.

CC The invention relates to a transgenic plant, comprising a genetic  
CC construct comprising a promoter operatively linked to a nucleic acid  
CC sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)  
CC family amino acid sequence (AAMS1651-AAMS1671). The transgenic plant has  
CC increased stress tolerance, especially to heat. The plant is a cereal,  
CC grass, ornamental plant, crop plant, food plant, oil-producing plant, a  
CC synthetic product-producing plant, an environmental waste absorbing  
CC plant, an alcohol plant, a medicinal plant, a recreational plant and/or  
CC an animal feed plant. In particular, the transgenic plant is cotton,  
CC canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or  
CC Arabidopsis thaliana. The plants may be used to produce animal feed,  
CC alcohol, crop, oil, medicine or a synthetic product. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained from GenBank using the Accession Number reference provided  
CC in the specification

XX Sequence 2992 BP; 847 A; 554 C; 777 G; 814 T; 0 U; 0 Other;

Query Match 10.4%; Score 32.4; DB 4; Length 2992;

Best Local Similarity 54.1%; Pred. No. 16; Mismatches 0; Indels 0; Gaps 0;

Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 82 AAAAGGTGTGGAAAAGTTGGATTGGTGGAAAAAGTTGCAATCTGACCATCTCTAGATC 141

DB 1176 AATACCGTGGAGAAATTGGAGGAAGATTGAAGAAGTTGATGGAGGAATTAACAAAGT 1235

QY 142 GTGGAAAAAGTTACGTTAGCGGTTTACGTACATATGTTGATTGTGAAAAAGAACGGA 201

DB 1236 GATGAATAATTTTATTATTGATGAAGTACATATCTTGTATGGGCGAGCAGCAGAA 1295

QY 202 GG 203

DB 1296 GG 1297

RESULT 14

ABN80109

ID ABN80109 standard; DNA; 6380 BP.

XX AC ABN80109;

XX 15-JUL-2002 (first entry)

XX Human chemically modified disease associated gene SEQ ID NO 126.

XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
KW antidiabetic; cytostatic; anticonvulsant; ds.

XX Homo sapiens.

OS Synthetic.

XX WO200200927-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007536.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130908/17.

XX Novel nucleic acid useful for diagnosis and therapy of diseases  
PT associated with development genes such as diabetes, comprises a sequence  
PT of a segment of chemically pretreated DNA of genes associated with  
PT development.

PS Claim 1; SEQ ID NO 126; 27pp; English.

XX The invention relates to a nucleic acid (I) comprising a sequence at  
CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
CC of genes associated with development selected from 87 genes listed in the  
CC specification such as ACCPN, ADFN, or AP1 and comprising one of 350  
CC sequences (ABN79984-ABN80333) or their complements. The invention is  
CC useful for the diagnosis or therapy of diseases associated with  
CC development genes, in particular disease related to homeobox containing  
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
CC associated with congenital heart disease, epilepsy, diseases related to  
CC histone deacetylation, Curtarino syndrome, diseases related with the  
CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
CC Oligomers specific to each of the genes are useful for detecting the  
CC methylation state of all CpG dinucleotides within the 350 sequences or  
CC (II) and their complementary sequences, as primer oligonucleotides for  
CC the amplification of the 350 sequences, (II) and/or their complements and  
CC as oligomer probes for detecting the cytosine methylation state and/or  
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this  
CC patent did not form part of the printed specification but is based on  
CC sequence information supplied to Derwent by the European Patent Office

XX Sequence 6380 BP; 1663 A; 193 C; 1758 G; 2765 T; 0 U; 1 Other;

Query Match 10.4%; Score 32.4; DB 6; Length 6380;

Best Local Similarity 62.2%; Pred. No. 22;

Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 46 AGTTGGTAAATACGAAAAAGAGATTTCATCCAGAAAAAGTGTGGAAAAAGTTGTGGAT 105

DB 5808 AGTTGATTAAGGGAATAAGGTTTTTAATAGAGATAGTTAAGTTTTTTGGAG 5867

QY 106 TCGGTGGAAAAAGTTGATCTG 127

DB 5868 TGGGTAGATAGAGTTTAATGTG 5889

RESULT 15

AAV52248/c

ID AAV52248 standard; DNA; 11303 BP.

XX AC AAV52248;

XX 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:115.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

XX WO9818931-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US019588.

XX 31-OCT-1996; 96US-0029960P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;

XX Dougherty BA;

XX WPI; 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae

XX polynucleotide sequences - useful in diagnostic kits and assays, and

XX pharmaceutical compositions and vaccines for Streptococcus pneumoniae.

XX Claim 1; Page 831-838; 1409pp; English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from *Streptococcus pneumoniae*. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the *S.pneumoniae* genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridize to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the *S. pneumoniae* genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the *S. pneumoniae* genome of commercial importance, or expression modulating fragments of the *S. pneumoniae* genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for *S. pneumoniae*.

Seq	Sequence	11303 BP;	3733 A;	2266 C;	1558 G;	3746 T;	0 U;	0 Other;
	Query Match	10.4%	Score 32.4;	DB 2;	Length 11303;			
	Best Local Similarity	51.4%;	Pred. No. 27;					
	Matches	75;	Conservative	0;	Mismatches	71;	Indels	0;
	Gaps	0;						
QY	62	AAAGAGATTTCATCATCCAGAAAAGGTGTGGAAAAGTTGTGGATTGGTTCGTCGAAAAAGTTC	121					
Db	6373	AGACAAGGTAAATTTACTTGGAAAAGATGGTTGAGAGTTTATCTGTGTGGTGAATCAAAATGA	6314					
QY	122	GATCTCACCATTCTCTAGATCGTGGAAAAAGTTCACGTTAGCGTTACGTCACATATGTGGA	181					
Db	6313	GTTTGAACAATCTTTACTTTATGGAATTAACAGCGCGAAGAAATTAAGTTCAGTTGTTTT	6254					
QY	182	TTCTGGAAAAAAGACGGAGGCATC	207					
Db	6253	TTCTGAAAAAATTCGAGTTCCGCATC	6228					

Search completed: July 3, 2005, 02:38:02  
Job time : 481.353 secs

**This Page Blank (usplc)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 03:19:18 ; Search time 640.235 Seconds  
(without alignments)  
3056.566 Million cell updates/sec

Title: US-10-814-858A-1  
Perfect score: 312  
Sequence: 1 gtcgaccatcatttgaagg.....tgacgcacttgacgtact 312

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCRS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:\*
- 20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*
- 21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*
- 22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:\*
- 24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	37.4	12.0	2515	13	US-10-027-632-59048 Sequence 59048, A
C 2	37.4	12.0	2515	17	US-10-027-632-59048 Sequence 59048, A
C 3	36.8	11.8	350764	13	US-10-087-192-1864 Sequence 1864, Ap
C 4	33.8	10.8	1491	19	US-10-437-963-52281 Sequence 52281, A
C 5	33.3	10.6	4074	17	US-10-320-797-210 Sequence 210, App
C 6	32.8	10.5	657	13	US-10-027-632-259460 Sequence 259460,
C 7	32.8	10.5	657	17	US-10-027-632-259460 Sequence 259460,

RESULT 1

US-10-027-632-59048/c  
; Sequence 59048, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 59048  
; LENGTH: 2515  
; TYPE: DNA

ALIGNMENTS

8	32.8	10.5	145806	20	US-10-719-993-6943	Sequence 6943, Ap
9	32.6	10.4	468	9	US-09-974-300-3550	Sequence 3550, Ap
10	32.6	10.4	585	19	US-10-767-701-29171	Sequence 29171, A
11	32.6	10.4	1163020	17	US-10-398-221-10	Sequence 10, Appl
12	32.6	10.4	3011208	17	US-10-398-221-2058	Sequence 2058, Ap
13	32.4	10.4	495	21	US-10-472-928-445	Sequence 445, App
14	32.4	10.4	2992	9	US-09-812-350-43	Sequence 43, Appl
C 15	32.4	10.4	11303	8	US-08-961-527-115	Sequence 115, App
C 16	32.4	10.4	11303	17	US-10-158-844-115	Sequence 11, Appl
C 17	32.4	10.4	65454	18	US-10-293-864-11	Sequence 50, Appl
C 18	32.4	10.4	83517	19	US-10-367-094-50	Sequence 17645, A
C 19	32.4	10.4	103660	21	US-10-741-600-17645	Sequence 4790, Ap
C 20	32.4	10.4	2162598	21	US-10-472-928-4979	Sequence 4979, Ap
C 21	32.2	10.3	370	19	US-10-767-701-4790	Sequence 236927,
C 22	32.2	10.3	625	13	US-10-027-632-236927	Sequence 236928,
C 23	32.2	10.3	625	17	US-10-027-632-236927	Sequence 236927,
C 24	32.2	10.3	625	17	US-10-027-632-236928	Sequence 236928,
C 25	32.2	10.3	625	17	US-10-027-632-236928	Sequence 12112, A
C 26	32	10.3	765	13	US-10-027-632-12112	Sequence 12113, A
C 27	32	10.3	765	17	US-10-027-632-12112	Sequence 12113, A
C 28	32	10.3	765	17	US-10-027-632-12112	Sequence 12113, A
C 29	32	10.3	765	17	US-10-027-632-12113	Sequence 1116, Ap
C 30	31.8	10.2	722	17	US-10-398-221-1116	Sequence 3638, Ap
C 31	31.8	10.2	2232	17	US-10-398-221-3638	Sequence 16091, A
C 32	31.8	10.2	3759	17	US-10-282-122A-16091	Sequence 45, Appl
C 33	31.6	10.1	592	9	US-09-770-152-45	Sequence 119, App
C 34	31.6	10.1	2526	10	US-09-934-455-119	Sequence 199, App
C 35	31.6	10.1	2526	15	US-10-278-536-199	Sequence 793, App
C 36	31.6	10.1	2526	17	US-10-225-086A-793	Sequence 145, App
C 37	31.6	10.1	2526	17	US-10-374-780A-145	Sequence 463, App
C 38	31.6	10.1	2526	18	US-10-412-699B-463	Sequence 1805, Ap
C 39	31.6	10.1	2526	18	US-10-412-699B-1805	Sequence 96, Appl
C 40	31.6	10.1	534	17	US-10-257-166-96	Sequence 177537,
C 41	31.4	10.1	443	13	US-10-027-632-177537	Sequence 177537,
C 42	31.4	10.1	443	17	US-10-027-632-177537	Sequence 1, Appli
C 43	31.4	10.1	3673778	16	US-10-312-841-1	Sequence 2105, Ap
C 44	31.4	10.0	9859	15	US-10-311-455-2105	Sequence 97972, A
C 45	31	9.9	783	18	US-10-424-599-97972	

```
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2515)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-59048

Query Match      12.0%; Score 37.4; DB 13; Length 2515;
Best Local Similarity 53.1%; Pred. No. 0.79;
Matches 77; Conservative 1; Mismatches 67; Indels 0; Gaps 0;

Qy 13 TTGAAAGGGCCTCGGTAATACCAATTTGGGAAAAGTTGGTAATACGGAAGAAAGATTC 72
Db 1051 TTAAGAGGGCCTAGATTATGATATAGGAAGGGCTAAGGCAATGTTGATAAAGCATGTTA 992

Qy 73 ATCATCCAGAAAAGGTGTGGAAGTTGTGGATTGGCTTGGAAAGTTGCATCTGACCAT 132
Db 991 ATGTGCTGGAAGCGTGTAAAGAAAGTAGTGTAGTTGATCATGAAGTGAGGCCCGGCAC 932

Qy 133 CTCTAGATCGTGAAGAAAGTTTCACG 157
Db 931 CCCAGGACAGGATGGAGGTCCACG 907
```

```
RESULT 2
US-10-027-632-59048/c
; Sequence 59048, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59048
; LENGTH: 2515
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2515)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-59048
```

```
Query Match      12.0%; Score 37.4; DB 17; Length 2515;
Best Local Similarity 53.1%; Pred. No. 0.79;
Matches 77; Conservative 1; Mismatches 67; Indels 0; Gaps 0;

Qy 13 TTGAAAGGGCCTCGGTAATACCAATTTGGGAAAAGTTGGTAATACGGAAGAAAGATTC 72
Db 1051 TTAAGAGGGCCTAGATTATGATATAGGAAGGGCTAAGGCAATGTTGATAAAGCATGTTA 992

Qy 73 ATCATCCAGAAAAGGTGTGGAAGTTGTGGATTGGCTTGGAAAGTTGCATCTGACCAT 132
Db 991 ATGTGCTGGAAGCGTGTAAAGAAAGTAGTGTAGTTGATCATGAAGTGAGGCCCGGCAC 932
```

```
Qy 133 CTCTAGATCGTGAAGAAAGTTTCACG 157
Db 931 CCCAGGACAGGATGGAGGTCCACG 907

RESULT 3
US-10-087-192-1864
; Sequence 1864, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1864
; LENGTH: 350764
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-1864

Query Match      11.8%; Score 36.8; DB 13; Length 350764;
Best Local Similarity 56.7%; Pred. No. 11;
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 70 TTCATCATCCAGAAAAGGTGTGGAAGTTGTGGATTGGCTTGGGAAAAGTTTCGATCTGAC 129
Db 176489 TTCTTTTAAAGAAATGATGCAAAAGCTGTTGCTTGGCGGGGTATTTTACATCAGAA 176548

Qy 130 CATCTCTAGATCGTGAAGAAAGTTTCACGTTAGCGCTTACGATCATATGTTGGAA 189
Db 176549 ATGCTTAAGAGATTGTAATAGGTTGGTGAAGAACTGAGGTACATTTAATGAATATGGA 176608
```

```
RESULT 4
US-10-437-963-52281
; Sequence 52281, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 52281
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_54592C.1
US-10-437-963-52281
```

```
Query Match      10.8%; Score 33.8; DB 19; Length 1491;
Best Local Similarity 53.4%; Pred. No. 8.3;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

```

, RESULT 6
, US-10-027-632-259460/c
, Sequence 259460, Application US/10027632
, Publication No. US20020198371A1
, GENERAL INFORMATION:
, APPLICANT: Wang, David G.
, TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
, TITLE OF INVENTION: Polymorphisms in the Human Genome
, FILE REFERENCES: 108827.129
, CURRENT APPLICATION NUMBER: US/10/027,632
, CURRENT FILING DATE: 2002-04-30
, PRIOR APPLICATION NUMBER: US 60/218,006
, PRIOR FILING DATE: 2000-07-12
, PRIOR APPLICATION NUMBER: US 60/198,676
, PRIOR FILING DATE: 2000-04-20
, PRIOR APPLICATION NUMBER: US 60/193,483
, PRIOR FILING DATE: 2000-03-29
, PRIOR APPLICATION NUMBER: US 60/185,218
, PRIOR FILING DATE: 2000-02-24
, PRIOR APPLICATION NUMBER: US 60/167,363
, PRIOR FILING DATE: 1999-11-23

```

RESULT 8  
US-10-719-993-6943

```
; Sequence 6943, Application US/10719993
; Publication No. US20040285849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6943
; LENGTH: 145806
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-6943

Query Match      10.5%; Score 32.8; DB 20; Length 145806;
Best Local Similarity 53.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 62;

QY 8 ATCAATTTGAAGGCGCTCGGTATACCATTTGTGGAAGAAAGTTGGTAATACGGAAAAAGAA 67
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 82601 ATTCATTGCAAGGTTGTTAAAGGACTAGAAAGCAAGAAAGGAATACAGGTTCAAAA 82660

QY 68 GATTATCATCCAGAAAAGTGTGGAAGAAAGTTGTGGATTCGCGTGGAAAAAGTTTCGATCTG 127
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 82661 GATTAAATATTCAGAAAGGTGTATTGCTCTGGGATGTAGAAGCAAGAAAAAATA 82720

QY 128 ACCATCTCTAGA 139
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 82721 ATATTACCCAGA 82732

RESULT 9
US-09-974-300-3550
; Sequence 3550, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3550
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-3550

Query Match      10.4%; Score 32.6; DB 9; Length 468;
Best Local Similarity 55.9%; Pred. No. 12;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 21 GCTCGGTAATACCATTTGGAAAAAGTTGGTAATACGGAAAAAGAAAGATTATCATCA 80
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191 GTCTCAACAGTCCGGCTTTACACGATTCAAAATCGGGAAGAAAGCGATTTTTTATCC 250

QY 81 GAAAGGTGTGGAAGAAAGTTGTGATTCGCGTGGAAAAAGTTTCGATCTGACCA 131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 GGAAGGGCTTGAACCTTTAGACAGTCCATGGAAAAAATGGGGAGCTGACAA 301

RESULT 10
US-10-767-701-29171

; Sequence 29171, Application US/10767701
; Publication No. US20040172864A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 29171
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 9296737
US-10-767-701-29171

Query Match      10.4%; Score 32.6; DB 19; Length 585;
Best Local Similarity 55.9%; Pred. No. 13;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 88 TGTGAAAAGTTGTGATTCGCGTGAAGAAAGTTTCGATCTGACCATCTCTAGATCGTGGAA 147
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 102 TGATCAAGAGTATACGATGGTTGAAACCGCGCGCAACACAGGCTCGACTTCGCGGAA 161

QY 148 AAGATTTCACGTAGCGCTTACGTACATATGTGGATTGTGAAAAAGAAAGAC 198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 162 AAGCTCGACGTGAGCGCATCGTAAAGCTATCGCGGAGGTCAAGAGNAGCC 212

RESULT 11
US-10-398-221-10
; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1163020
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

Query Match      10.4%; Score 32.6; DB 17; Length 1163020;
Best Local Similarity 51.7%; Pred. No. 3.6e+02;
Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 62 AAAGAAGATTTCATCATCCAGAAAGGTGTGAAAAAGTTGTGGATTTCGCGTGGAAAAAGTTTC 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61761 AAATCAGGTTTATAAATCAAAAATCGGGTGAAGAGATGAATAAGAAGATGAAGAAAGCTT 61820

QY 122 GATCTGACCATCTCTAGATCGTGGAAAAAGTTTCAGTTAGCGCTTACGTACATATGTGGA 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61821 TTTAGTATTACTTCTGTTGTTGTTAGTATTATTCGTTAGTGGTTGTTATCAATTGTTGCA 61880

QY 182 TTGTGAAAAAGAACGAGGAGC 204
```



```
Db 61881 GAGTCAAAACGGCGGAAGCGGC 61903
|||||
RESULT 12
US-10-398-221-2058
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
US-10-398-221-2058
Query Match 10.4%; Score 32.6; DB 17; Length 3011208;
Best Local Similarity 51.7%; Pred. No. 4.9e+02;
Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 62 AAGAAGATTCATCCAGAAAGGTGTGGAAGTTGTGGATTCGGTGGAAAAAGTTTC 121
Db 146002 AAATGAGGTTTATAAATCAAAATCGGGTGAAGAGATGAATAAGAAGATGAAGAAAGCTT 146061
QY 122 GATCTGACCATCTCTAGATCGTGAAGAAAGTTTCAGTTAGCGTTAGGTACATATGTGGA 181
Db 146062 TTTAGTATTACTTCTGTGTGTGTGTAGTATTATCGTTAGTGTGTATCAATTTGGTGA 146121
QY 182 TTGTGGAAGAAAGAGCGGAGGC 204
Db 146122 GAGTCAAAACGGCGGAAGCGGC 146144
RESULT 13
US-10-472-928-445
; Sequence 445, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: THE INSTITUTE FOR GENOMIC RESEARCH
; FILE REFERENCE: P026926W0
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 445
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-472-928-445
Query Match 10.4%; Score 32.4; DB 21; Length 495;
Best Local Similarity 51.4%; Pred. No. 14;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 62 AAGAAGATTCATCCAGAAAGGTGTGGAAGTTGTGGATTCGGTGGAAAAAGTTTC 121
Db 105 AGACAAGGTAATTTACTTGAAGAGATGGTTGAGAGTTTATCTGTGGTGAATCAATGA 164

QY 122 GATCTGACCATCTCTAGATCGTGAAGAAAGTTTCAGTTAGCGTTAGGTACATATGTGGA 181
Db 165 GTTTGAACAATCTTTACTTTATGGAATTAACAGCGCGAAGAATTAAGTTCAGTTGTTTT 224
QY 182 TTGTGGAAGAAAGAGCGGAGGCATC 207
Db 225 TTCTGAAAAAATTCAGTTCCTCCGCATC 250

RESULT 14
US-09-812-350-43
; Sequence 43, Application US/09812350
; Publication No. US20020053097A1
; GENERAL INFORMATION:
; APPLICANT: Lindquist, Susan
; APPLICANT: Queitsch, Christine
; APPLICANT: Vierling, Elizabeth
; TITLE OF INVENTION: Transgenic Plants Containing Heat Shock Protein
; FILE REFERENCE: P01979US2
; CURRENT APPLICATION NUMBER: US/09/812,350
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/190,769
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: US 60/198,116
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
; LENGTH: 2992
; TYPE: DNA
; ORGANISM: Spinacia oleracea
US-09-812-350-43
Query Match 10.4%; Score 32.4; DB 9; Length 2992;
Best Local Similarity 54.1%; Pred. No. 31;
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 82 AAAAGGTGTGGAAGTTGTGGATTCGGTGGAAAAAGTTTCGATCTGACCATCTCTAGATC 141
Db 1176 AAATACCGTGGAGATTTCAGGAAGATTGAAGAAGTTGATGGAGGAATTAACAAAGT 1235
QY 142 GTGGAAAAAGTTTCAGTTAGCGTTAGGTACATATGTGGATTTGTGAAAAAGACGGA 201
Db 1236 GATGAATTAATTTATTATTATGATGAAGTACATCTTTGATTTGGGCGAGCAGCAGAA 1295
QY 202 GG 203
Db 1296 GG 1297

RESULT 15
US-08-961-527-115/C
; Sequence 115, Application US/08961527
; Publication No. US20020032323A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
```

```
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-115

Query Match      10.4%; Score 32.4; DB 8; Length 11303;
Best Local Similarity 51.4%; Pred. No. 55;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY      62 AAAGAAGATTTCATCCAGAAAGGTGTGGAAGTTGTGGATTGCGTGGAAAAAGTTC 121
Db      6373 AGACAAGGTAAATTACTTGAAGAAGATGGTTGAGAGTTTATCTGTGGTGAATCAATGA 6314

QY      122 GATCTGACCATCTTAGATCGTGGAAAAAGTTCCACGTTAGCGCTTACGTACATATGTGGA 181
Db      6313 GTTTGAACAATCTTACTTTATGGAATTAACACAGCGGAAGATTAAAGTTCAGTTGTTTT 6254

QY      182 TTGTGGAAGAAAGACGGAGGCATC 207
Db      6253 TTCTGAAAAAATTGCAGTTCGCCATC 6228
```

Search completed: July 3, 2005, 07:33:13  
Job time : 652.235 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 02:01:47 ; Search time 3925.41 Seconds  
(without alignments)

3025.428 Million cell updates/sec

Title: US-10-814-858A-1

Perfect score: 312

Sequence: 1 gtcgaccatcattgaaag.....tgacgcacttgacgtact 312

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_hic.\*

4: gb\_est3.\*

5: gb\_est4.\*

6: gb\_est5.\*

7: gb\_est6.\*

8: gb\_gss1.\*

9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.8	12.4	777	8	BZ096570 CH230-236
2	38	12.2	1007	9	AL076752 Drosophil
3	37.8	12.1	1101	9	AL073282 Drosophil
4	37.2	11.9	754	8	BZ063129 11d86503.
5	36.6	11.7	925	9	AL074475 Drosophil
6	36.4	11.7	765	9	CC807465 ZMMB8C051
7	36.4	11.7	2707	3	AK048004 Mus muscu
8	36.2	11.6	633	7	CF979725 tg80c09.Y
9	36	11.5	368	8	BZ840661 CH240.261
10	36	11.5	576	9	CE352559 tigr-gss-
11	36	11.5	960	9	AL066226 Drosophil
12	36	11.5	1070	9	AG120841 Pan trogl
13	35.8	11.5	367	8	BZ852956 CH240.217
14	35.6	11.4	921	9	AG561995 Mus muscu
15	35.6	11.4	1101	9	CNS0106X.
16	35.4	11.3	507	9	CR498396 Medicago
17	35.4	11.3	508	8	AQ439862 HS 5059.A
18	35.4	11.3	512	9	EX185081 Danio rer
19	35.4	11.3	513	8	AZ570160 271PvB08
20	35.4	11.3	585	9	BX198554 Danio rer
21	35.4	11.3	1449	9	AG391030 Mus muscu
22	35.2	11.3	375	9	CE224908 tigr-gss-
23	35.2	11.3	392	9	CE224907 tigr-gss-
24	35.2	11.3	567	4	BI667582 603292418

c	25	35.2	11.3	751	8	BZ074256	BZ074256 lkh06g12.
	26	35.2	11.3	797	9	CNS003P8	AL064634 Drosophil
	27	35.2	11.3	924	9	CNS0056T	AL057103 Drosophil
	28	35	11.2	1002	9	CL948189	CL948189 OslFESB003
	29	34.8	11.2	621	6	CB435223	CB435223 615287.MA
	30	34.8	11.2	642	4	BG538051	BG538051 602563558
	31	34.8	11.1	945	7	CF584573	CF584573 AGENCOURT
	32	34.6	11.1	1100	9	CNS00G3S	AL071398 Drosophil
	33	34.4	11.0	492	1	AV675381	AV675381 AV675381
	34	34.4	11.0	611	5	BW045972	BW045972 BW045972
	35	34.4	11.0	630	5	BW264718	BW264718 BW264718
	36	34.4	11.0	641	5	BW337262	BW337262 BW337262
	37	34.4	11.0	647	5	BW052958	BW052958 BW052958
	38	34.4	11.0	651	5	BW030523	BW030523 BW030523
	39	34.4	11.0	681	5	BW301819	BW301819 BW301819
	40	34.4	11.0	700	5	BW486169	BW486169 BW486169
	41	34.4	11.0	714	7	CF368490	CF368490 853144.MA
	42	34.4	11.0	755	5	BW299977	BW299977 BW299977
c	43	34.4	11.0	759	5	BW424685	BW424685 BW424685
	44	34.4	11.0	769	5	BW043106	BW043106 BW043106
	45	34.4	11.0	789	1	AL666171	AL666171 AL666171

## ALIGNMENTS

RESULT 1  
BZ096570  
LOCUS BZ096570 777 bp DNA linear GSS 10-OCT-2002  
DEFINITION CH230-236C5-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-236C5, genomic survey sequence.  
ACCESSION BZ096570  
VERSION BZ096570.1 GI:23737454  
KEYWORDS GSS.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 777)  
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Teegaye,G., Gear,K.,  
Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,  
Riggs,F., de Jong,P. and Fraser,C.M.  
Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
Unpublished (1999)  
JOURNAL Other GSSs: CH230-236C5.TJ  
COMMENT Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(http://www.chori.org/bacpac/rat230.htm). For BAC library  
availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources  
(http://www.chori.org/bacpac/or ering information.htm). BAC end  
page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
Plate: 236 row: C column: 5  
Seq primer: T7  
Class: BAC ends.

## FEATURES

Location/Qualifiers  
1..777  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SsNHd/MCw"  
/db\_xref="taxon:10116"  
/clone="CH230-236C5"  
/sex="Female"  
/cell\_type="Brain"  
/clone\_lib="CHORI-230 Segment 1"  
/note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;

```
CHORI-230 Rat (BN/SenHed/MCW) BAC library produced by
Pieter de Jong"

ORIGIN
Query Match          12.4%; Score 38.8; DB 8; Length 777;
Best Local Similarity 51.1%; Pred. No. 1.8;
Matches 91; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 43 AAAAGTTGGTAATACGGAAGAAAGATTTCATCATCCAGAAAAGGTGTGGAAGTTGTG 102
    |||||
Db 193 AGAAGAAGCAAAATTTTATAAGAATATTAAAACAAAACAAATACAGAAATATAGAT 252
    |||||

QY 103 GATTGCGTGAAAAGTTTCGATCTGACCATCTCTAGATCGTGGAAGTTTCACGTTAGC 162
    |||||
Db 253 AAATCAAGTCAAAATTCGACTACTAACAAATAACACAGCAAGGAAAGAAATCAATTAAGC 312
    |||||

QY 163 GCTTACGTCATCATATGTTGGTGGAAAAGAGAGCGGAGGCATCGTGGAAAAGAA 220
    |||||
Db 313 ACTTATGGGTTAAGTAGAATAAGAAATAGAAATGTAATGAAGGATGGAAAAGGA 370
    |||||

RESULT 2
CNS00JOV/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR38B02 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION
AL076752.1 GI:4956128
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1007)
AUTHORS
TITLE
JOURNAL
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
source
1..1007
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR38B02"
/clone_lib="RPCI-98"
/notes="end : TET3"

ORIGIN
Query Match          12.2%; Score 38; DB 9; Length 1007;
Best Local Similarity 39.7%; Pred. No. 3.4;
Matches 50; Conservative 29; Mismatches 47; Indels 0; Gaps 0;

QY 35 ATTGTGGAAAAGTTGGTAATACGGAAGAAAGAGATTTCATCCAGAAAAGGTGTGGAA 94
    ||:|
Db 654 AWDADAADAATAAATARAATAATAKAAAKAAAAAAATKTGGTARAATAWAKTGRKKDAA 595
    ||:|
```

```
QY 95 AAGTTGTGGATTGCGTGGAAAAGTTGATCTGACCATCTCTAGATCGTGGAAGATTC 154
    ||:|
Db 594 AAKKTTTGTGKKGKTKAATAWATTKGTGTARGADARTNKGAGAAKAAARAAAGGK 535
    ||:|

QY 155 ACGTTA 160
    ||:|
Db 534 AXGGDA 529
    ||:|

RESULT 3
CNS00HBP
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR34A09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION
AL073282.1 GI:4953062
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
TITLE
JOURNAL
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR34A09"
/clone_lib="RPCI-98"
/notes="end : TET3"

ORIGIN
Query Match          12.1%; Score 37.8; DB 9; Length 1101;
Best Local Similarity 34.2%; Pred. No. 4;
Matches 96; Conservative 46; Mismatches 139; Indels 0; Gaps 0;

QY 15 GAAAGGCGCTCGGTAATACCATTTGGAAAAAGTTGGTAATACGGAAGAAAGATTCAT 74
    |||||
Db 741 GAAAGGTTTTGKMWTTATCAATGYASAAGAGGTGTTAKDAYCAAGAGGGMTTHTMAR 800
    |||||

QY 75 CATCCAGAAAAGGTGTGGAAGTTGTGGATTGCGTGGAAAAGTTTCGATCTGACCATCT 134
    |||||
Db 801 TTWTHAHMGATGTGCAWAWMDTTCTMWCARTCTGGGAAWAWATTTKGARTMTYCAAT 860
    |||||

QY 135 CTAGATCGTGGAAAAGTTTCACGTTAGCGCTTACGTACATATGTGCGATTGTGGAAGA 194
    |||||
Db 861 TTWTCGKSGGGGMMWTGGAAGAGCCCCCTTTTAWAAAAAARAAAGMTATAWAWMA 920
    |||||

QY 195 AGACGGAGGCATCGGTGGAAAAGAGCTTGTACGCTGTACGCTAGCATAGATAC 254
    |||||
```

Db 921 AAAAAANGWATGTTATGGGAWAANAATWTGTGTATGTATGTCGCCGGCTCTTGWTAT 980

Qy 255 ACGTGCACGGCTCCACTTGACGCAATTAAGCACAATGA 295

Db 981 GGGCTKAHAGAGTATGCGMAATAWAAGGGGKYMTSWG 1021

RESULT 4

BZ063129

LOCUS

DEFINITION 754 bp DNA linear GSS 10-OCT-2002

ACCESSION BZ063129

VERSION BZ063129.1 GI:23673631

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea

Brassica oleracea

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 754)

Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T., Nash, W., Rabinowicz, P.D. and Wilson, R.K.

Whole genome shotgun reads from Brassica oleracea

Unpublished (2002)

Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@wustl.edu

Plate: lld86 row: g column: 03

Seq primer: -21UpPof forward

Class: shotgun

High quality sequence start: 16

High quality sequence stop: 551.

FEATURES

source

1..754

Location/Qualifiers

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/db\_xref="taxon:3712"

/clone\_lib="B.oleracea002"

/notes="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea T0100DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

ORIGIN

Query Match 11.9%; Score 37.2; DB 8; Length 754;

Best Local Similarity 53.4%; Pred. No. 5.5;

Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 59 GAAAAAGAGATTCATCATCCAGAAAGGTGTGAAAAGTTGGATTGGTGGAAAAAG 118

Db 462 GAAAGGCGAGCGAAGTAGAAAAACATTCAGTCAATTTGATTCTCTATAAAATAT 521

Qy 119 TTCATCTGACCATCTCTACATCGTGAAAAAGTTCACGTTAGCGTTACGTACATATGT 178

Db 522 TTGGTATATTTTCATGTAAAGTGTGAAAAAGTGTCTTAAACGGAACCAATATTT 581

Qy 179 GGAATTGGAAAAAGACGAGGC 204

Db 582 GGATTACAAAAAAGGGATTATGC 607

RESULT 5

CNS001BN

LOCUS

DEFINITION 925 bp DNA linear GSS 03-JUN-1999

Brassica oleracea

Drosophila melanogaster genome survey sequence T7 end of BAC: BACR36N21 of RCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL074475

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 925)

Genoscope.

Direct Submission

TITLE

AUTHORS

JOURNAL

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Ooeogawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

source

1..925

Location/Qualifiers

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone\_lib="BACR36N21"

/clone\_lib="RCI-98"

/note="end : T7"

ORIGIN

Query Match 11.7%; Score 36.6; DB 9; Length 925;

Best Local Similarity 15.3%; Pred. No. 8.6;

Matches 33; Conservative 97; Mismatches 86; Indels 0; Gaps 0;

Qy 19 GGGCTTCGTAATACCATTTGGAAAAAGTTGGTAATACGAAAAAGAGATTTCATC 78

Db 668 GCGMGGGGGVMKKKKKXMDAMKMKMKMKMKMKKKKKKKKKKKKKKKKKKKKK 727

Qy 79 CAGAAAAGTGTGAAAAAGTTGTGGATTTCGGTGGAAAAAGTTGCGATCTCTAG 138

Db 728 RAAAKADKKKKKKAADAAADDDKAKKKKOHAMMMMMMMMTKTKTKKOHMMMTTOMMTK 787

Qy 139 ATCGTGGAAAAAGTTTCACGTTAGCGTTACGTACATATGTGGATTGTGAAAAAGAGAC 198

Db 788 KGAAGAARAKKKAANAKAAKAKKKKAKKKKAKKKKAKKKKAKKKKAKKADAAA 847

Qy 199 GGAGGATCGGTGGAAAAAGAGCTTTGTACGCTGTA 234

Db 848 AKKKKKKADKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 883

RESULT 6

CC807465

LOCUS

DEFINITION 765 bp DNA linear GSS 14-JUL-2003

ZMMBBc0514M03f ZMMBBc Zea mays genomic clone ZMMBBc0514M03 5', genomic survey sequence.

CC807465

ACCESSION CC807465.1 GI:32638048

VERSION

KEYWORDS

SOURCE

ORGANISM

Zea mays

Zea mays

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 765)

AUTHORS Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,  
Rouyard,K., Fuks,G., Yu,Y., Wing,R. and Messing,J.  
TITLE Sequencing of the maize genome at PGIR (2003b)  
JOURNAL Unpublished (2003)  
COMMENT Contact: Bharti,A.K.  
Dr Joachim Messing's lab  
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers  
University  
190 Frelinghuysen Road, Piscataway, NJ 08854, USA  
Tel: 732 445 3801  
Fax: 732 445 5735  
Email: bharti@waksman.rutgers.edu  
Seq primer: T7  
Class: BAC ends  
High quality sequence start: 401.  
Location/Qualifiers  
1. .765  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ZMBBc0514M03"  
/lab\_host="E. coli DH10B"  
/clone\_lib="ZMBBc"  
/note="Vector: pTARBAC1.3; Site\_1: BamHI; Site\_2: BamHI"

ORIGIN  
Query Match 11.7%; Score 36.4; DB 9; Length 765;  
Best Local Similarity 50.6%; Pred. No. 9.5;  
Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 81 GAAAGGTGTGGAAGTTGTGTGATTCGCGTGGAAAAAGTTGCGATCTGACCATCTTAGAT 140  
|||||  
Db 491 GAAGACAGGCGGAGAATTGGGGATCGAGAGGCTAAACCTAAACTAGACTAGAACTACTC 550  
|||||

QY 141 CGTGGAAAAGTTCCAGTTAGCGCTTACGTACATATGTGGATGTGGAAAAAGAACCG 200  
|||||

Db 551 CTAAGATAAAATCGGAAGTAGAAGTTGTATGTGATTCGATTTGATTACAAATCGGCCGT 610  
|||||

QY 201 AGGCATCGGTGGAAAAAGAGCTTGTACGCTGACGCTGACGATAGATAGATAC 254  
|||||

Db 611 AGACCTCTTTATATAGAGAGGGGGGCTGGACCTTTACAGCTGATTC 664  
|||||

RESULT 7  
AK048004  
LOCUS  
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:Cl30029F03 product:unclassified, full insert sequence.  
ACCESSION AK048004  
VERSION AK048004.1 GI:26092584  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
PUBMED 11042159  
REFERENCE  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
PUBMED 11076861  
4  
REFERENCE  
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
REFERENCE  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
REFERENCE  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/.  
Location/Qualifiers  
1. .2707  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strains="CS7BL/6J"  
/db\_xref="FANTOM,DB:Cl30029F03"  
/db\_xref="taxon:10090"  
/clone="Cl30029F03"  
/tissue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="16 days embryo"  
1. .2707  
/note="unclassified"

misc\_feature  
ORIGIN  
Query Match 11.7%; Score 36.4; DB 3; Length 2707;  
Best Local Similarity 52.7%; Pred. No. 13;  
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 149 AAGTTTCAGCTTAGCGCTTACGTACATATGTGATTTGTGAAAAAGAACGAGGAGCATCG 208  
|||||  
Db 2145 AAATTCCTTTTCTCTTACACATATGTGATGTGTGCACACACACACACAGG 2204  
|||||

Qy 209 GTGAAAAGAAGCTTGTACGCTGACGATAGATAGATACACGTCACGCGTCC 268  
 |||||  
 Db 2205 GTGCATACACATATGTGTATGATATACACACACACACACACACACACAC 2264  
 |||||  
 Qy 269 ACTTGACGCACAATGACGCACAATGACGC 298  
 |||||  
 Db 2265 ACACACACACACTCACACACACACACAC 2294  
 |||||

RESULT 8  
 LOCUS CF979725 633 bp mRNA linear EST 25-NOV-2003  
 DEFINITION r80609.y1 Meloidogyne hapla female SL1 pGEM Meloidogyne hapla cDNA  
 5', similar to TR:Q5XW8 Q9XW8 Y37D8A.14 PROTEIN. [1] ;, mRNA  
 sequence.  
 CF979725  
 CF979725.1 GI:38512774  
 EST.  
 Meloidogyne hapla  
 Meloidogyne hapla  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
 Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.  
 1 (bases 1 to 633)  
 McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,  
 Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,  
 Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,  
 Taagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,  
 Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,  
 Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,  
 McCann,R., Waterston,R. and Wilson,R.  
 The Washington Univ. Nematode EST Project, 1999  
 Unpublished (1999)  
 Contact: McCarter JP

TITLE JOURNAL  
 COMMENT  
 The Washington Univ. Nematode EST Project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 Cloned unidirectionally. Poly (A)+ RNA was concentrated and  
 purified using Dynabeads (Dyna) and mRNA eluted for 1st strand  
 synthesis. 1st strand cDNA created using MMLV RT(PowerScript,  
 Clontech) and primed with oligo(dT) with XhoI site (primer  
 CDSIII/3-XhoI) and 5' SMART anchor added using chimeric DNA-RNA  
 oligo (SMART-NotI-r-GGG). SMART-NotI-r-GGG was used in case a PCR  
 using the SMART template was also desired. 15 PCR cycles were done  
 using 1st strand and primers specific to SL1 leader sequence  
 (SL1-NotI) and 3' end (XhoI-No-dT). dscDNA was digested using  
 XhoI/NotI, fractionated on Chroma-spin 400 columns (Clontech) and  
 ligated to digested pGEM-11zf(+) plasmid. Chemically competent  
 cells were used as host. This library was provided by David Bird's  
 lab (david.bird@ncsu.edu), of North Carolina State University.  
 Seq primer: Sp6.

FEATURES  
 source  
 1. .633  
 Location/Qualifiers  
 /organism="Meloidogyne hapla"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:6305"  
 /sex="female"  
 /tissue\_type="whole organism"  
 /dev\_stage="adult"  
 /lab\_host="DH108"  
 /clone\_lib="Meloidogyne hapla female SL1 pGEM"  
 /note="Vector: pGEM-11zf(+) (Promega); Site 1: XhoI;  
 Site 2: NotI; Cloned unidirectionally. Poly (A)+ RNA was  
 concentrated and purified using Dynabeads (Dyna) and mRNA  
 eluted for 1st strand synthesis. 1st strand cDNA created  
 using MMLV RT(PowerScript, Clontech) and primed with  
 oligo(dT) with XhoI site (primer CDSIII/3-XhoI) and 5'  
 SMART anchor added using chimeric DNA-RNA oligo  
 (SMART-NotI-r-GGG). SMART-NotI-r-GGG was used in case a  
 PCR using the SMART template was also desired. 15 PCR

cycles were done using 1st strand and primers specific to  
 SL1 leader sequence (SL1-NotI) and 3' end (XhoI-No-dT).  
 dscDNA was digested using XhoI/NotI, fractionated on  
 Chroma-spin 400 columns (Clontech) and ligated to digested  
 pGEM-11zf(+) plasmid. Chemically competent cells were  
 used as host. This library was provided by David Bird's  
 lab (david.bird@ncsu.edu), of North Carolina State  
 University."

## ORIGIN

Query Match 11.8%; Score 36.2; DB 7; Length 633;  
 Best Local Similarity 49.7%; Pred. No. 10;  
 Matches 92; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
 Qy 11 ATTGGAAGGGCTCCGTAATACCATTTGCGAAAAAGTTGGTAATACGGAAGAGAT 70  
 |||||  
 Db 449 ATGACAAATGCTCGATATGAACATTTTCATAGAGTTGTAGAGAAATAAATAGAA 508  
 |||||  
 Qy 71 TCATCATCCAGAAAGGTGTGAAAAAGTTGTGATTCGCTGGAAAAAGTTCTGATCTGACC 130  
 |||||  
 Db 509 AGAAGTGTGTAATATTTGTAGGGATATTTGTAATGATTTAAAAAATTAATTTTCC 568  
 |||||  
 Qy 131 ATCTAGATCGTGGAAAAAGTTTCACGTTAGCCCTACGTACATATGTGATTTGTGAAA 190  
 |||||  
 Db 569 CTCTTTAAAGTAAAAAATATTTTGAAGAATTTCTGATTAATAATTTATTTTAAAAA 628  
 |||||  
 Qy 191 AAGAA 195  
 |||||  
 Db 629 AAAAA 633

## RESULT 9

LOCUS BZ840661 368 bp DNA linear GSS 18-MAR-2003  
 DEFINITION CH240\_261B20.TJ CHORI-240 Bos taurus genomic clone CH240\_261B20,  
 genomic survey sequence.  
 ACCESSION BZ840661  
 VERSION BZ840661.1 GI:29068020  
 KEYWORDS GSS.  
 SOURCE Bos taurus (cow)  
 ORGANISM

REFERENCE  
 AUTHORS Zhao,S., Shetty,J., Shateman,S., Tsengaye,G., Geer,K.,  
 Shvartsbeyn,A., Gebregorgis,E., Chen,D., Riggs,F., de Jong,P.,  
 Crawford,A.M. and McEwan,J.C.  
 Bovine BAC End Sequences from Library CHORI-240  
 Unpublished (2003)  
 TITLE JOURNAL  
 COMMENT Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@igr.org

Clones are derived from the bovine BAC library CHORI-240  
 (http://www.chori.org/bacpac/bovine240.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/ordering/information.htm). This work  
 was undertaken as part of the International Bovine BAC Mapping  
 Consortium (IBBMC) by AgResearch Ltd., New Zealand and The  
 Institute of Genomic Research (TIGR), USA.  
 Plate: 261 row: B column: 20  
 Seq primer: Sp6  
 Class: BAC ends.

FEATURES  
 source

1. .368  
 Location/Qualifiers  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /strain="breed: Hereford"





```

Db      793 GGNMMGCGVGGNG 779

RESULT 12
AG120841/c
LOCUS   AG120841          1070 bp      DNA      linear      GSS 04-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-129L09.F, genomic survey sequence.
ACCESSION AG120841
VERSION   AG120841.1  GI:16650006
KEYWORDS  GSS.
SOURCE   Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes

REFERENCE
AUTHORS  Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE    BAC end sequences of library PTB
JOURNAL  Unpublished
AUTHORS  Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE    Direct Submission
JOURNAL  Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
          Tel:81-45-503-9111 Fax:81-45-503-9170)
COMMENT  Clones are derived from the chimpanzee BAC library PTB This BAC end
          was generated during the R&D process and may have higher chance of
          clone tracking errors.
PRIMERS

Sequencing: -21M13
LIBRARY
Vector   : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
source   Location/Qualifiers
1..1070
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-129L09.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
Query Match      11.5%; Score 36; DB 9; Length 1070;
Best Local Similarity 48.5%; Pred. No. 13;
Matches 99; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY  44  AAGTTGGTAAATACGGAAAAGAGATTTCATCTCGATCGTGGAAAAGTTTCAGTTAGCG 103
      |||
Db   1070 AATAAGGGAAGGGGAAAAGCGGAGGAATAGGAGAAAGAGGAGGAGGAGATAGATGGGA 1011
      |||

QY  104  ATTGCGTGGAAAAGTTTCGATCTGACCATCTCTAGATCGTGGAAAAGTTTCAGTTAGCG 163
      |||
Db   1010 ATAGGGAGAAAAGAGAGATAGAGATAAATAAAGGAGGAGGAGGAGGAGGT 951

QY  164  CTTACGTACATATGTGGATTGTGGAAAAGACGCGAGGCATCGGTGGAAAAGAGGCT 223
      |||
Db   950  CGAAAGAGAGAGGTGGAGAGGAGAGAAAGAAAGAGAGAGTGAAGAGGAGGAGGAGGAGGAG 891

QY  224  TGTACGCTGACGTGACGATAGA 247
      |||
Db   890  GGGAGGGGACAGAAAAGAGAGAGA 867

RESULT 13
BZ852956
LOCUS   BZ852956          367 bp      DNA      linear      GSS 18-MAR-2003
DEFINITION CH240_217K24.TV CHORI-240 Bos taurus genomic clone CH240_217K24,
          genomic survey sequence.
ACCESSION BZ852956
VERSION   BZ852956.1  GI:48322693
KEYWORDS  GSS.

```

```

ACCESSION BZ852956
VERSION   BZ852956.1  GI:29080361
KEYWORDS  GSS.
SOURCE   Bos taurus (cow)
ORGANISM Bos taurus

REFERENCE
AUTHORS  Zhao,S., Shetty,J., Shatsman,S., Teegaye,G., Geer,K.,
          Shvartsbeyn,A., Gebregeorgis,E., Chen,D., Riggs,F., de Jong,P.,
          Crawford,A.M. and McSwan,J.C.
TITLE    Bovine BAC End Sequences from Library CHORI-240
JOURNAL  Unpublished (2003)
COMMENT  Contact: Shaying Zhao
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: szhao@tigr.org
          Clones are derived from the bovine BAC library CHORI-240
          (http://www.chori.org/bacpac/bovine240.htm). For BAC library
          availability, please contact Pieter de Jong (pdejong@mail.cho.org).
          Clones may be purchased from BACPAC Resources
          (http://www.chori.org/bacpac/ordering/information.htm). This work
          was undertaken as part of the International Bovine BAC Mapping
          Consortium (IBBMC) by AgResearch Ltd., New Zealand and The
          Institute of Genomic Research (TIGR), USA.
          Plate: 217 row: K column: 24
          Seq primer: T7
          Class: BAC ends.
          Location/Qualifiers
1..367
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_217K24"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site 1: Mbol; Site 2: Mbol;
          Hereford bull Li Domino 99375; CHORI-240 Bovine BAC
          library (Male) produced by Pieter de Jong"

ORIGIN
Query Match      11.5%; Score 35.8; DB 8; Length 367;
Best Local Similarity 50.9%; Pred. No. 12;
Matches 85; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY  5  ACCATCATTTGAAGGCGCTCGGTAAATACCATTTGTGAAAAAGTTTGGTAATACGAAAAA 64
      |||
Db   181  ACAAAAAATTTAAAACTGCAGCTTCATATCATTTAAAAACAAAGAAACTGAACATAAAAA 240
      |||

QY  65  GAAGATTTCATCCAGAAAAGGTGTGGAAGATTTCGGATTCCGTTGGAAAAAGTTTCGAT 124
      |||
Db   241  GATGAAAAAATATCCAAAAGGCAAAATATGAAATATATGCTATTAAATAAGAGACATTATTA 300
      |||

QY  125  CTGACCATCTCTAGATCGTGGAAAAGTTTCAGCTTAGCGCTTACGTA 171
      |||
Db   301  CAGGACATCTTACATTATAGAAAACCTTGAACCTCTATTGATAATTA 347
      |||

RESULT 14
AG561995
LOCUS   AG561995          921 bp      DNA      linear      GSS 05-JUN-2004
DEFINITION Mus musculus molossinus DNA, clone: MSMg01-481K04.T7, genomic survey
          sequence.
ACCESSION AG561995
VERSION   AG561995.1  GI:48322693
KEYWORDS  GSS.

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: July 3, 2005, 00:53:27 ; Search time 197.647 Seconds  
(without alignments)  
3893.639 Million cell updates/sec  
Title: US-10-814-858A-2  
Perfect score: 130  
Sequence: 1 tcactatataggaagttc.....ttacatctagataaacaatg 130  
Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 4390206 seqs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Query		Match		Length		DB		ID		Description	
Result No.	Score	Match	Length	DB	ID	DB	ID	DB	ID	DB	ID	DB	ID	DB	ID
1	130	100.0	452	3	AAA28449	AAA28449	Artificia	AAA28449	Artificia	AAA28449	Artificia	AAA28449	Artificia	AAA28449	Artificia
2	89	68.5	89	3	AAA28445	AAA28445	Synthetic	AAA28445	Synthetic	AAA28445	Synthetic	AAA28445	Synthetic	AAA28445	Synthetic
3	63.6	48.9	3212	5	AAF30968	AAF30968	Aspergill	AAF30968	Aspergill	AAF30968	Aspergill	AAF30968	Aspergill	AAF30968	Aspergill
4	63.6	48.9	3288	5	AAF30970	AAF30970	Aspergill	AAF30970	Aspergill	AAF30970	Aspergill	AAF30970	Aspergill	AAF30970	Aspergill
5	63.6	48.9	3329	5	AAF30969	AAF30969	Aspergill	AAF30969	Aspergill	AAF30969	Aspergill	AAF30969	Aspergill	AAF30969	Aspergill
6	61	46.9	1006	2	AAx81436	AAx81436	Arabidops	AAx81436	Arabidops	AAx81436	Arabidops	AAx81436	Arabidops	AAx81436	Arabidops
7	60.4	46.5	1138	2	AAT85665	AAT85665	pWJBI DNA	AAT85665	pWJBI DNA	AAT85665	pWJBI DNA	AAT85665	pWJBI DNA	AAT85665	pWJBI DNA
8	60.4	46.5	1138	2	AAT59713	AAT59713	Plasmid P	AAT59713	Plasmid P	AAT59713	Plasmid P	AAT59713	Plasmid P	AAT59713	Plasmid P
9	60.4	46.5	1138	2	AAx78852	AAx78852	Plasmid P	AAx78852	Plasmid P	AAx78852	Plasmid P	AAx78852	Plasmid P	AAx78852	Plasmid P
10	59.6	45.8	157	6	AAD4423	AAD4423	PGI482 co	AAD4423	PGI482 co	AAD4423	PGI482 co	AAD4423	PGI482 co	AAD4423	PGI482 co
11	58.2	44.8	2200	6	ABL58982	ABL58982	HPV16-L1	ABL58982	HPV16-L1	ABL58982	HPV16-L1	ABL58982	HPV16-L1	ABL58982	HPV16-L1
12	56.2	43.2	5033	2	AAQ74695	AAQ74695	pPH11406	AAQ74695	pPH11406	AAQ74695	pPH11406	AAQ74695	pPH11406	AAQ74695	pPH11406
13	52	40.0	6395	2	AAQ95155	AAQ95155	TMV-based	AAQ95155	TMV-based	AAQ95155	TMV-based	AAQ95155	TMV-based	AAQ95155	TMV-based
14	52	40.0	6395	2	AAZ20642	AAZ20642	TMV-based	AAZ20642	TMV-based	AAZ20642	TMV-based	AAZ20642	TMV-based	AAZ20642	TMV-based
15	52	40.0	6395	3	AAC62369	AAC62369	cDNA sequ	AAC62369	cDNA sequ	AAC62369	cDNA sequ	AAC62369	cDNA sequ	AAC62369	cDNA sequ
16	52	40.0	6395	4	AAF82330	AAF82330	Wild-type	AAF82330	Wild-type	AAF82330	Wild-type	AAF82330	Wild-type	AAF82330	Wild-type
17	52	40.0	6425	2	AAZ20644	AAZ20644	TMV-based	AAZ20644	TMV-based	AAZ20644	TMV-based	AAZ20644	TMV-based	AAZ20644	TMV-based
18	52	40.0	6425	4	AAF82332	AAF82332	Tobacco m	AAF82332	Tobacco m	AAF82332	Tobacco m	AAF82332	Tobacco m	AAF82332	Tobacco m
19	52	40.0	6439	2	AAZ20643	AAZ20643	TMV-based	AAZ20643	TMV-based	AAZ20643	TMV-based	AAZ20643	TMV-based	AAZ20643	TMV-based
20	52	40.0	6439	4	AAF82331	AAF82331	Tobacco m	AAF82331	Tobacco m	AAF82331	Tobacco m	AAF82331	Tobacco m	AAF82331	Tobacco m

21	52	40.0	6446	2	AAZ20646	AAZ20646	TMV-based	AAZ20646	TMV-based	AAZ20646	TMV-based	AAZ20646	TMV-based	AAZ20646	TMV-based
22	52	40.0	6446	4	AAF82334	AAF82334	Tobacco m	AAF82334	Tobacco m	AAF82334	Tobacco m	AAF82334	Tobacco m	AAF82334	Tobacco m
23	52	40.0	6475	2	AAZ20645	AAZ20645	TMV-based	AAZ20645	TMV-based	AAZ20645	TMV-based	AAZ20645	TMV-based	AAZ20645	TMV-based
24	52	40.0	6475	4	AAF82333	AAF82333	Tobacco m	AAF82333	Tobacco m	AAF82333	Tobacco m	AAF82333	Tobacco m	AAF82333	Tobacco m
25	52	40.0	7926	3	AAZ89190	AAZ89190	Plasmid T	AAZ89190	Plasmid T	AAZ89190	Plasmid T	AAZ89190	Plasmid T	AAZ89190	Plasmid T
26	52	40.0	10600	11	ADM68432	ADM68432	Celery CE	ADM68432	Celery CE	ADM68432	Celery CE	ADM68432	Celery CE	ADM68432	Celery CE
27	52	40.0	10600	12	ADP26585	ADP26585	TMV CEL I	ADP26585	TMV CEL I	ADP26585	TMV CEL I	ADP26585	TMV CEL I	ADP26585	TMV CEL I
28	52	40.0	10600	13	ADQ88407	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I
29	52	40.0	10600	13	ADQ88407	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I
30	52	40.0	10624	11	ADM68433	ADM68433	Celery CE	ADM68433	Celery CE	ADM68433	Celery CE	ADM68433	Celery CE	ADM68433	Celery CE
31	52	40.0	10624	12	ADP26586	ADP26586	TMV CEL I	ADP26586	TMV CEL I	ADP26586	TMV CEL I	ADP26586	TMV CEL I	ADP26586	TMV CEL I
32	52	40.0	10624	13	ADQ88408	ADQ88408	TMV CEL I	ADQ88408	TMV CEL I	ADQ88408	TMV CEL I	ADQ88408	TMV CEL I	ADQ88408	TMV CEL I
33	52	40.0	10624	13	ADQ88407	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I	ADQ88407	TMV CEL I
34	52	40.0	11222	12	ADN97500	ADN97500	Artificia	ADN97500	Artificia	ADN97500	Artificia	ADN97500	Artificia	ADN97500	Artificia
35	52	40.0	11641	6	AAZ45229	AAZ45229	TTODA xGA	AAZ45229	TTODA xGA	AAZ45229	TTODA xGA	AAZ45229	TTODA xGA	AAZ45229	TTODA xGA
36	52	40.0	11641	6	AAZ45229	AAZ45229	TTODA xGA	AAZ45229	TTODA xGA	AAZ45229	TTODA xGA	AAZ45229	TTODA xGA	AAZ45229	TTODA xGA
37	52	40.0	11641	10	ADQ84772	ADQ84772	Tobacco m	ADQ84772	Tobacco m	ADQ84772	Tobacco m	ADQ84772	Tobacco m	ADQ84772	Tobacco m
38	52	40.0	11641	12	ADQ88298	ADQ88298	TTODA xGA	ADQ88298	TTODA xGA	ADQ88298	TTODA xGA	ADQ88298	TTODA xGA	ADQ88298	TTODA xGA
39	52	40.0	11641	12	ADQ88298	ADQ88298	TTODA xGA	ADQ88298	TTODA xGA	ADQ88298	TTODA xGA	ADQ88298	TTODA xGA	ADQ88298	TTODA xGA
40	51.8	39.8	1825	2	AAQ67663	AAQ67663	TMV repli	AAQ67663	TMV repli	AAQ67663	TMV repli	AAQ67663	TMV repli	AAQ67663	TMV repli
41	51.8	39.8	1825	2	AAV54825	AAV54825	Replicon	AAV54825	Replicon	AAV54825	Replicon	AAV54825	Replicon	AAV54825	Replicon
42	51.8	39.8	7688	5	ADQ20293	ADQ20293	Tobacco m	ADQ20293	Tobacco m	ADQ20293	Tobacco m	ADQ20293	Tobacco m	ADQ20293	Tobacco m
43	51.8	39.8	7688	6	ADQ24477	ADQ24477	Tobacco m	ADQ24477	Tobacco m	ADQ24477	Tobacco m	ADQ24477	Tobacco m	ADQ24477	Tobacco m
44	51.8	39.8	7688	8	ADAI5016	ADAI5016	DNA trans	ADAI5016	DNA trans	ADAI5016	DNA trans	ADAI5016	DNA trans	ADAI5016	DNA trans
45	51.8	39.8	7688	10	ADB83324	ADB83324	DNA trans	ADB83324	DNA trans	ADB83324	DNA trans	ADB83324	DNA trans	ADB83324	DNA trans

ALIGNMENTS

RESULT 1  
AAA28449  
ID AAA28449 standard; DNA; 452 BP.  
XX  
AC AAA28449;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Artificial synthetic promoter for high level plant gene expression.  
XX  
KW Artificial promoter; strong; weak; transgene expression; plant; ss.  
XX  
OS Synthetic.  
XX  
FH Key  
FT misc\_signal 1..120 Location/Qualifiers  
FT misc\_signal 121..142 /tag= a  
FT misc\_signal 143..220 /notes= "domain II"  
FT misc\_signal 221..266 /tag= b  
FT misc\_signal 267..307 /notes= "domain III"  
FT misc\_signal 308..324 /tag= c  
FT misc\_signal 325..351 /notes= "Minimal domain Ib"  
FT misc\_signal 352..452 /tag= d  
FT TATA\_signal 353..374 /notes= "Minimal domain Ia"  
FT TATA\_signal 375..452 /tag= e  
FT 5'UTR 453..452 /notes= "plant consensus TATA signal"  
FT EP1002869-Al.  
PD 24-MAY-2000.  
XX

```

PF 25-FEB-1999; 99EP-00301419.
XX
PR 09-NOV-1998; 98IN-DE003322.
XX
XX (COUL ) CSIR COUNCIL SCI IND RES.
XX
PI Tuli R, Sawant SV, Singh PK, Gupta SK;
XX
XX WPI; 2000-341712/30.
XX
DR
XX
XX New chemically synthesized artificial promoter, useful high level
PT expression of transgenes in different organisms.
XX
XX Claim 13; Page 11; 40pp; English.
XX
XX Chemically synthesized artificial promoters are new and comprise a DNA
CC sequence designed for a targeted level and pattern of gene expression by
CC strategically putting together several signature sequences identified by
CC sequence alignment and statistical analysis of a large database
CC constructed for this purpose. A method for chemically synthesizing an
CC artificial promoter for expressing genes at a desired level in different
CC organisms is also claimed. The high level expression in a plant using
CC such an artificial promoter (e.g. AAA28449) can be measured comprising
CC polyethylene glycol (PEG) mediated transformation of plant protoplasts as
CC well as biolistic mediated transformation of plant tissues including
CC root, stem, intact leaf tissue followed by transient GUS assay to compare
CC with a natural CaMV 35S promoter showing the desired level of activity.
CC The promoter is useful for high level expression of transgenes in
CC different organisms and for testing high level gene expression in plants
CC (claimed). The promoter is biologically active and is efficient and can
CC be synthesized to express in even the most complex organisms
XX
XX Sequence 452 BP; 157 A; 88 C; 95 G; 112 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 130; DB 3; Length 452;
Best Local Similarity 100.0%; Pred. No. 8.5e-22;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCACATATATAGGAAGTTTCATTTTGGAAATGGACACGTTGTTCATTTCTCAACAA 60
Db 313 TCACATATATAGGAAGTTTCATTTTGGAAATGGACACGTTGTTCATTTCTCAACAA 372
QY 61 TTACCAACACACACACACACACACACACACACACACACACACACACACACACACAC 120
Db 373 TTACCAACACACACACACACACACACACACACACACACACACACACACACACACAC 432
QY 121 ATAAACAATG 130
Db 433 ATAAACAATG 442
RESULT 2
AAA28445
ID AAA28445 standard; DNA; 89 BP.
XX
AC AAA28445;
XX
XX 29-AUG-2000 (first entry)
XX
XX Synthetic promoter 5' untranslated leader sequence.
XX
KW Artificial promoter; 5' untranslated leader sequence;
KW transgene expression; plant; ss.
XX
OS Synthetic.
XX
XX EP1002869-A1.
XX
XX 24-MAY-2000.
XX
XX 25-FEB-1999; 99EP-00301419.
XX
XX 09-NOV-1998; 98IN-DE003322.
XX

```

```

XX (COUL ) CSIR COUNCIL SCI IND RES.
XX
XX Tuli R, Sawant SV, Singh PK, Gupta SK;
XX
XX WPI; 2000-341712/30.
XX
XX New chemically synthesized artificial promoter, useful high level
PT expression of transgenes in different organisms.
XX
XX Claim 9; Page 23; 40pp; English.
XX
XX This comprises a 5' untranslated leader (UL) sequence (+1 to +89) which
CC contains a translational enhancer "CAA" type region. The 5' UL of highly
CC expressed genes varies from 75-90 nucleotides, while that of the lowly
CC expressed genes varies from 100-300 nucleotides. Chemically synthesized
CC artificial promoters are new and comprise a DNA sequence designed for a
CC targeted level and pattern of gene expression by strategically putting
CC together several signature sequences identified by sequence alignment and
CC statistical analysis of a large database constructed for this purpose. A
CC method for chemically synthesizing an artificial promoter for expressing
CC genes at a desired level in different organisms is also claimed. The high
CC level expression in a plant using such an artificial promoter (e.g.
CC AAA28449) can be measured comprising polyethylene glycol (PEG) mediated
CC transformation of plant protoplasts as well as biolistic mediated
CC transformation of plant tissues including root, stem, intact leaf tissue
CC followed by transient GUS assay to compare with a natural CaMV 35S
CC promoter showing the desired level of activity. The promoter is useful
CC for high level expression of transgenes in different organisms and for
CC testing high level gene expression in plants (claimed). The promoter is
CC biologically active and is efficient and can be synthesized to express in
CC even the most complex organisms
XX
XX Sequence 89 BP; 41 A; 20 C; 4 G; 24 T; 0 U; 0 Other;
SQ
Query Match 68.5%; Score 89; DB 3; Length 89;
Best Local Similarity 100.0%; Pred. No. 4.3e-12;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 39 ACGTGTGTGTCATTTCTCAACATTTACCAACACACACACACACACACACACATTTATCAAT 98
Db 1 ACGTGTGTGTCATTTCTCAACATTTACCAACACACACACACACACACACATTTATCAAT 60
QY 99 TACTATTACATTTACATCTAGATAACA 127
Db 61 TACTATTACATTTACATCTAGATAACA 89
RESULT 3
AAF30968
ID AAF30968 standard; cDNA; 3212 BP.
XX
XX AAF30968;
XX
XX 23-JUL-2001 (first entry)
XX
XX Aspergillus niger beta-glucosidase BGL1 cDNA in plasmid pJDB1.
XX
KW Beta-glucosidase; BGL1; aroma; alcoholic beverage; ethanol;
KW transgenic plant; tobacco; pJDB1; ss.
XX
XX Aspergillus niger.
XX
XX Key Location/Qualifiers
FH mat_peptide 391..2916
FT /*tag= a
XX
XX WO200136586-A2.
XX
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-IL000758.
XX

```

```

FT sig_peptide 391..507
FT /*tag= b
FT /note= "Cell signal peptide"
FT mat_peptide 508..3039
FT /*tag= c
FT /note= "encodes BGL1-HDEL fusion"
XX
XX WO200136586-A2.
XX
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-IL000758.
XX
XX 19-NOV-1999; 99US-00443338.
XX
XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Shu W, Marton I, Siegel DL, Ben-Ami B, Dekel M, Shoseyov O;
XX
XX WPI; 2001-355624/37.
XX
XX P-PSDB; AAB82328.
XX
XX New nucleic acid encoding beta-galactosidase from Aspergillus niger,
XX useful e.g. for increasing aroma compound formation during alcoholic
XX fermentation.
XX
XX Disclosure; Page 101-103; 106pp; English.
XX
XX The present sequence is that of a cDNA construct in plasmid pJDCB1T
XX encoding a polypeptide (see AAB82328) comprising the Cell signal peptide,
XX followed by the mature region of Aspergillus niger B1 (GMI CC 324626)
XX beta-glucosidase BGL1 and the endoplasmic reticulum (ER) retaining
XX peptide HDEL. The cDNA is placed under the control of the 35S promoter
XX and nos terminator. The gene cassette in pJDCB1T was inserted into binary
XX vector pBINplus.. Disarmed Agrobacterium tumefaciens LB4404 was
XX transformed with the resulting vector and used to transform tobacco cv.
XX Samson plants. The gene expression cassette provided for accumulation of
XX BGL1 in the ER. BGL1 polynucleotides, polypeptides, host cells and
XX transgenic plants are useful for fermentation, particularly to increase
XX the level of aroma compounds in alcoholic drinks, as well as other
XX fermentation products of plant material, hydrolysing cellobiose and thus
XX increasing the level of fermentable glucose, increasing the production of
XX alcohol, such as ethanol, from plant material, increasing the aroma
XX released from a plant or a plant product, and increasing the hydrolysis
XX or transglycosylation of glycosides. Unlike previously known beta-
XX glucosidases, BGL1 is very stable over a wide range of temperatures and
XX pH, e.g. 68% retention of activity after 1 hour at 60 degree C
XX
XX Sequence 3288 BP; 782 A; 873 C; 876 G; 757 T; 0 U; 0 Other;
XX
XX Query Match 48.9%; Score 63.6; DB 5; Length 3288;
XX Best Local Similarity 87.7%; Pred. No. 6.2e-06;
XX Matches 93; Conservative 0; Mismatches 9; Indels 4; Gaps 2
XX
XX QY 12 AGGAAGTTCATTTCATTGGAAATGGACA--CGTGTTCATTTCTCAACAATTACCAACA 69
XX |||||
XX Db 274 AGGAAGTTCATTTCATTGGAGAGGACAGCTCTTCGATGCTTCAACAATTACCAACA 333
XX |||||
XX QY 70 ACAACAAACAAACAAACAACTTATACAAATTACTATTACAAATTACA 115
XX |||||
XX Db 334 ACAACAAACAAACAAACAACT--TACAAATTACTATTACAAATTACA 377
XX |||||
XX
XX RESULT 5
XX AAF30969.
XX ID AAF30969 standard; cDNA; 3329 BP.
XX
XX AAF30969;
XX
XX DT 23-JUL-2001 (first entry)
XX
XX Aspergillus niger beta-glucosidase BGL1 cDNA in plasmid pJDCB1.
XX

```

KW Beta-glucosidase; BGL1; aroma; alcoholic beverage; ethanol;  
 KW transgenic plant; tobacco; pUDCB1; ss.

XX  
 OS Aspergillus niger.  
 OS Unidentified.  
 OS Chimeric.

XX Key Location/Qualifiers  
 FT CDS 391..3033  
 FT /\*tag= a  
 FT sig\_peptide 391..507  
 FT /\*tag= b  
 FT /\*note= "Cell signal peptide"  
 FT mat\_peptide 508..3030  
 FT /\*tag= c  
 FT /\*note= "BGL1 mature polypeptide"

XX WO200136586-A2.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000WO-IL000758.

XX 19-NOV-1999; 99US-00443338.

XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

XX Shu W, Marton I, Siegel DL, Ben-Ami B, Dekel M, Shoseyov O;

XX WPI: 2001-355624/37.  
 DR P-PSDB; AAB82327.

XX New nucleic acid encoding beta-galactosidase from Aspergillus niger,  
 PT useful e.g. for increasing aroma compound formation during alcoholic  
 PT fermentation.

XX Disclosure; Page 97-99; 106pp; English.

XX The present sequence is that of a cDNA construct in plasmid pUDCB1  
 CC encoding a polypeptide (see AAB82327) comprising the Cell signal peptide  
 CC and the mature region of Aspergillus niger B1 (CWI CC 324626) beta-  
 CC glucosidase BGL1. The cDNA is placed under the control of the 35S  
 CC promoter and nos terminator. The gene cassette in pUDCB1 was inserted  
 CC into binary vector pBINplus. Disarmed Agrobacterium tumefaciens LB4404  
 CC was transformed with the resulting vector and used to transform tobacco  
 CC cv. Samson plants. The gene expression cassette provided for secretion of  
 CC BGL1 into the apoplast. BGL1 polynucleotides, polypeptides, host cells  
 CC and transgenic plants are useful for fermentation, particularly to  
 CC increase the level of aroma compounds in alcoholic drinks, as well as  
 CC other fermentation products of plant material, hydrolysing cellobiose and  
 CC thus increasing the level of fermentable glucose, increasing the  
 CC production of alcohol, such as ethanol, from plant material, increasing  
 CC the aroma released from a plant or a plant product, and increasing the  
 CC hydrolysis or transglycosylation of glycosides. Unlike previously known  
 CC beta-glucosidases, BGL1 is very stable over a wide range of temperatures  
 CC and pH, e.g. 68% retention of activity after 1 hour at 60 degree C

XX Sequence 3329 BP; 790 A; 882 C; 883 G; 774 T; 0 U; 0 Other;

Query Match 48.9%; Score 63.6; DB 5; Length 3329;  
 Best Local Similarity 87.7%; Pred. No. 6.2e-06;  
 Matches 93; Conservative 0; Mismatches 9; Indels 4; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAAATGGACA--CGTGTGTGTCATTTCTCAACAATTACCAACA 69

DB 274 AGGAAGTTCATTTCATTTGGAGAGGACAGCTTCTTGAGATCCTTCAACAATTACCAACA 333

QY 70 ACACAAACACACAAACAATTTACAAATTTACAAATTACA 115

DB 334 ACAACAAACAAACAACAACAT--TACAATTAATTTACAAATTACA 377

RESULT 6

AAx81436

ID AAX81436 standard; DNA; 1006 BP.

XX

AC AAX81436;

XX 25-AUG-1999 (first entry)

XX Arabidopsis thaliana promoter fragment produced by PCR.

XX Plant promoter; minimal promoter; transcription activating element;  
 KW gene expression; transgenic plant; ss.

XX Synthetic.

OS Arabidopsis thaliana.

XX WO9931258-A1.

XX 24-JUN-1999.

XX 10-DEC-1998; 98WO-EP008162.

XX 12-DEC-1997; 97EP-00203912.

XX (MOGE-) MOGEN INT NV.

XX Stuiver MH, Sijbolts FH;

XX WPI; 1999-395188/33.

XX Novel constitutive plant promoters.

XX Example 10; Page 36-37; 44pp; English.

XX The specification describes plant promoters, where elements from one  
 CC promoter which are responsible for specific expression are combined with  
 CC elements from a second promoter which are responsible for a complementary  
 CC expression pattern. The novel promoter then shows expression in the  
 CC tissues and developmental stages which form the expression pattern of  
 CC both promoters. The plant promoter is characterized in that it comprises  
 CC a minimal promoter and transcription activating elements from a set of  
 CC promoters, the elements having a complementary pattern and level of  
 CC transcription in a plant. The promoters of the invention are used to  
 CC express genes in transgenic plants. The present sequence was used in the  
 CC course of the invention

XX Sequence 1006 BP; 320 A; 182 C; 183 G; 321 T; 0 U; 0 Other;

Query Match 46.9%; Score 61; DB 2; Length 1006;  
 Best Local Similarity 88.3%; Pred. No. 2.4e-05;  
 Matches 91; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAAATGGACACGTCGTGTGTCATTTCTCAACAATTACCAACAAC 71

DB 906 AGGAAGTTCATTTCATTTGGAGAGGACAGGTAT----TTTACAAACAATTACCAACAAC 960

QY 72 AACAAACAAACAACAACATTTACAAATTTACAAATTTACAAATTAC 114

DB 961 AACAAACAAACAACAACAT--TACAATTAATTTACAAATTAC 1001

RESULT 7

AAT85665

ID AAT85665 standard; DNA; 1138 BP.

XX

AC AAT85665;

XX 21-NOV-1997 (first entry)

XX pmJB1 DNA sequence.

XX expression cassette; inducible promoter; alcA; alcR; alda; regulator;  
 KW alcohol dehydrogenase; herbicide resistance gene; glyphosate;  
 KW N-phosphonomethyl-glycine; glyphosate oxidase; GOX; EPSPS;

KW 5-enol-pyruvylshikimate 3-phosphate CP4; chloroplast transit sequence 1;  
 KW Arabidopsis; RUBISCO; CPT1; dicotyledon; monocotyledon; crop; ss.  
 XX Synthetic.  
 OS Synthetic.  
 XX WO9706269-A1.  
 PN 20-FEB-1997.  
 XX 02-AUG-1996; 96WO-GB001883.  
 PF 03-AUG-1995; 95GB-00015941.  
 XX (ZENE ) ZENECA LTD.  
 PA Jenson I;  
 PI WPI; 1997-154273/14.  
 XX Chemically inducible cassette for expressing herbicide resistance gene in  
 PT plants - and derived plants, partic. for resistance to glyphosate, avoids  
 PT constitutive expression and minimises development of herbicide tolerant  
 PT weeds.  
 XX Example 3; Fig 8; 59pp; English.

XX A chemically inducible plant gene expression cassette which comprises an  
 CC inducible promoter linked to a gene (I) that confers resistance to a  
 CC herbicide, is claimed. (I) imparts resistance to N-phosphonomethyl-  
 CC glycine (glyphosate) or its salts, but may also be a gene for resistance  
 CC to, e.g. chloroacetanilides, glufosinate, sulphonyl ureas, imidazolonolones,  
 CC etc. The inducible promoter (e.g. alcA, alcR, alcA or other alcR-induced  
 CC gene promoter) is operatively linked to an alcR regulator sequence.  
 CC Induced expression of (I) avoids the risk that constitutive expression  
 CC may interfere with plant development; allows volunteer plants to be  
 CC controlled by herbicide applied without inducer and minimises the chance  
 CC that herbicide-resistant strains of weeds will arise. The expression  
 CC cassette is strictly controlled and suitable for general use in plants  
 CC (both mono- and dicotyledons). The alcA/alcR gene switch was exemplified  
 CC with genes conferring resistance to glyphosate. The switch was used to  
 CC drive inducible expression of glyphosate oxidase (GOX) in plants.  
 CC Switchable GOX was expressed alone or in conjunction with constitutive  
 CC expression of 5-enol-pyruvylshikimate 3-phosphate (EPSPS) CP4. Constructs  
 CC were optimised for expression in mono- and dicotyledonous crop species.  
 CC The present sequence is that of plasmid pWJBI, based on pIBT211  
 CC containing the CamV 35 promoter with duplicated enhancer linked to the  
 CC tobacco mosaic virus translational enhancer sequence replacing the  
 CC tobacco etch virus 5' non-translated leader, and terminated with the  
 CC nopaline synthase poly(A) signal (nos). pWJBI was used in construction of  
 CC dicot vectors

SQ Sequence 1138 BP; 370 A; 253 C; 234 G; 281 T; 0 U; 0 Other;

Query Match 46.5%; Score 60.4; DB 2; Length 1138;  
 Best Local Similarity 85.8%; Pred. No. 3.3e-05;  
 Matches 91; Conservative 0; Mismatches 11; Indels 4; Gaps 2;

QY 12 AGGAAGTTCATTTTCATTTGGAGACACGTGTGTCATTCTCAACAATTACCAACAAC 71  
 DB 743 AGGAAGTTCATTTTCATTTGGAGAGG--ACCTCGAGTATTTTACAAACAATTACCAACAAC 800  
 QY 72 ACAAACAACAACAACAATTATCAATTTACTATTACAAATTACATC 117  
 DB 801 ACAAACAACAACAACAAT--TACAATTACTATTACAAATTACACC 844

RESULT 8  
 AAT59713  
 ID. AAT59713 standard; DNA; 1138 BP.  
 XX  
 AC AAT59713;  
 XX

DT 17-OCT-2003 (revised)  
 DT 12-MAY-1997 (first entry)  
 XX Plasmid pWJBI used for use in delta-endotoxin expression.  
 DE Gene expression cassette; promoter; alcR regulator; insecticide;  
 XX CryIA(c); CryV; crystal protein; delta-endotoxin; Bacillus thuringiensis;  
 KW Lepidoptera; insect resistance; transgenic plant; pWJBI; crop protection;  
 KW biological control; alcR; gene switch; ds.  
 XX Brassica oleracea var. botrytis; mosaic virus.  
 OS Nicotiana tabacum; mosaic virus.  
 OS Agrobacterium tumefaciens.  
 OS Chimeric.  
 XX Key Location/Qualifiers  
 FT TATA\_signal 737..743  
 FT /\*tag= c

XX WO9706268-A2.  
 PN 20-FEB-1997.  
 PD 29-JUL-1996; 96WO-GB001846.  
 XX 08-AUG-1995; 95GB-00016241.  
 XX (ZENE ) ZENECA LTD.

PA Jenson I, Paine JAM;  
 PI WPI; 1997-154272/14.  
 DR Chemically inducible expression cassette - contains inducible promoter  
 XX activated by alcR regulator in presence of alcohol or ketone inducer,  
 PT used for insecticide production in plants.  
 PT Example 4; Fig 11; 52pp; English.

XX Plasmid pWJBI (AAT59713) is based on pIBT211 and contains the cauliflower  
 CC mosaic virus 35S promoter with duplicated enhancer linked to the tobacco  
 CC mosaic virus (TMV) translational enhancer sequence and a nopaline-  
 CC synthase 3' gene region. Incorporation of an endotoxin CryV (see also  
 CC AAT59702) or CryIA(c) (see also AAT59712) gene into the plasmid  
 CC downstream of the TMV enhancer provides vectors for the constitutive or  
 CC inducible expression of CryV or CryIA(c). Transformation of plants  
 CC allows the generation of transgenic plants resistant to insect pests.  
 CC (Updated on 17-OCT-2003 to standardise OS field)

SQ Sequence 1138 BP; 369 A; 254 C; 234 G; 281 T; 0 U; 0 Other;

Query Match 46.5%; Score 60.4; DB 2; Length 1138;  
 Best Local Similarity 85.8%; Pred. No. 3.3e-05;  
 Matches 91; Conservative 0; Mismatches 11; Indels 4; Gaps 2;

QY 12 AGGAAGTTCATTTTCATTTGGAGACACGTGTGTCATTCTCAACAATTACCAACAAC 71  
 DB 743 AGGAAGTTCATTTTCATTTGGAGAGG--ACCTCGAGTATTTTACAAACAATTACCAACAAC 800  
 QY 72 ACAAACAACAACAACAATTATCAATTTACTATTACAAATTACATC 117  
 DB 801 ACAAACAACAACAACAAT--TACAATTACTATTACAAATTACACC 844

RESULT 9  
 AAX78852  
 ID AAX78852 standard; cDNA; 1138 BP.  
 XX  
 AC AAX78852;  
 XX 07-SEP-1999 (first entry)  
 DT Plasmid pWJBI EcoRI-HindIII cDNA fragment.  
 XX

```

XX Paraquat; antibody; light chain; herbicide; resistant; crop plant;
KW weed control; tolerant; diquat; photosynthesis inhibitor; photosystem I;
KW free radical; lipid peroxidation; electron transport; photosystem II;
KW vacuole; cell surface; cytotoxic; sensitive; ss.
XX
OS Synthetic.
XX
XX WO9932630-A1.
XX
XX 01-JUL-1999.
XX
XX 15-DEC-1998; 98WO-GB003760.
XX
XX 19-DEC-1997; 97GB-00026955.
XX
XX (ZENE ) ZENECA LTD.
XX
XX Holt DC, Jones PG;
XX
XX WPI; 1999-405173/34.
XX
XX Herbicide binding proteins and related polynucleotides.
XX
XX Disclosure; Page 46-47; 60pp; English.
XX
XX This invention describes a novel herbicide binding protein which can
CC confer herbicide resistance activity. Crop plants, such as soybean,
CC cotton, tobacco, sugarbeet, oilseed rape, canola, flax, sunflower,
CC potato, tomato, alfalfa, lettuce, maize, wheat, sorghum, rye, bananas,
CC barley, oat, turf grass, forage grass, sugar cane, pea, field bean, rice,
CC pine, poplar, apple, grape, citrus or nut plants, transformed with a
CC herbicide binding protein gene are resistant to the herbicide. Hence,
CC weeds can be selectively controlled in a field of the transformed crops.
CC The plants are substantially resistant or tolerant to herbicides, such as
CC paraquat or diquat, that inhibit photosynthesis by accepting electrons
CC from photosystem I thus generating free radicals which cause lipid
CC peroxidation or by blocking electron transport in photosystem II. The
CC herbicide binding proteins advantageously sequester the herbicide, e.g.
CC at the cell surface or in the vacuoles of a treated plant. Sequestration
CC at the cell surface prevents the entry of the herbicide into the cell so
CC that the herbicide cannot reach its intracellular target and exert any
CC significant cytotoxic effect. The herbicide binding protein inhibits the
CC mobility of the herbicide from the application site to the whole plant
CC preventing the herbicide reaching particularly sensitive organs.
CC Additionally, tolerant plants can be produced against herbicides that
CC have more than one target site
XX
XX Sequence 1138 BP; 370 A; 255 C; 233 G; 280 T; 0 U; 0 Other;
XX
Query Match 46.5%; Score 60.4; DB 2; Length 1138;
Best Local Similarity 85.8%; Pred. No. 3.3e-05;
Matches 91; Conservative 0; Mismatches 11; Indels 4; Gaps 2;
XX
QY 12 AGGAAGTTCATTTCGAAATGGACAGTGTTGTCATTCTCACAAATTACCAAC 71
Db 743 AGGAAGTTCATTTCATTGGAGAGG--ACCTCAGATGTTTACAAATTTACCAAC 800
XX
QY 72 AACAAACAACAACAACATTATACAAATTACTATTACAATTACATC 117
Db 801 AACAAACAACAACAACAT--TACAAATTACTATTACAAATTACACC 844
XX
RESULT 10
AAD44423
ID AAD44423 standard; DNA; 157 BP.
XX
XX AAD44423;
XX
XX 29-AUG-2003 (revised)
DT 13-DEC-2002 (first entry)
XX
XX pGL482 construct comprising 35S-PCISV promoter DNA.

```

```

XX Cauliflower mosaic virus 35S RNA; 35S; promoter; transgenic plant; OPR;
KW peanut chlorotic streak caulimovirus full length transcript promoter;
KW PCISV; 12-oxophytodienoic acid-100, 11-reductase gene promoter; SAG12;
KW transgenic; CaMV; chimeric; ds.
XX
XX Peanut; chlorotic streak caulimovirus.
OS
OS Cauliflower mosaic virus.
OS Chimeric.
XX
XX US6388170-B1.
XX
XX 14-MAY-2002.
XX
XX 07-APR-2000; 2000US-00545244.
XX
XX 07-APR-2000; 2000US-00545244.
XX
XX (KENT ) UNIV KENTUCKY RES FOUND.
XX
XX Gan S, Xie M, He Y;
XX
XX WPI; 2002-442888/47.
XX
XX Bi-directional promoter with common promoter linked in opposite
PT orientation to minimal promoter, useful to direct expression of
PT polynucleotide which confers agronomically significant trait to
PT transgenic plant.
XX
XX Disclosure; Col 13-16; 16pp; English.
XX
XX The invention relates to an artificial nucleic acid construct comprising
CC a bi-directional promoter having a minimal promoter operably linked in
CC opposite orientation 5' to a common promoter, where each promoter is
CC either cauliflower mosaic virus (CaMV) 35S RNA promoter (35S), peanut
CC chlorotic streak caulimovirus full length transcript promoter (PCISV),
CC Arabidopsis 12-oxophytodienoic acid-100, 11-reductase gene promoter (OPR)
CC or SAG12, and where each end of the bi-directional promoter is operably
CC linked to a polynucleotide encoding a polypeptide. The construct is used
CC to produce a transgenic plant which has an agronomically-significant
CC trait. The present sequence is pGL482 construct comprising 35S-PCISV
CC promoter DNA. (Updated on 29-AUG-2003 to standardise OS field)
XX
XX Sequence 157 BP; 57 A; 38 C; 22 G; 40 T; 0 U; 0 Other;
XX
Query Match 45.8%; Score 59.6; DB 6; Length 157;
Best Local Similarity 87.3%; Pred. No. 4.4e-05;
Matches 89; Conservative 0; Mismatches 9; Indels 4; Gaps 2;
XX
QY 12 AGGAAGTTCATTTCATTGGAAATGGACA--CGTGTGTCATTCTCAACAATTACCAACA 69
Db 41 AGGAAGTTCATTTCATTGGAGAGGACAGGCTCTTGAGATCCTTCAACAATTACCAACA 100
XX
QY 70 ACAACAACAACAACAACATTATACAAATTACTATTACAAT 111
Db 101 ACAACAACAACAACAACAT--TACAATTACTATTACAAT 140
XX
RESULT 11
ABL58982
ID ABL58982 standard; DNA; 2200 BP.
XX
XX ABL58982;
XX
XX 22-JUL-2002 (first entry)
DT
XX HPV16-L1 encoding DNA 2.
DE
XX Human papilloma virus; HPV; HPV16-L1; HPV16-L2; vaccine; cervical cancer;
KW HPV16-L1; cytostatic; virucide; gene; ds.
XX
XX Human papillomavirus.
OS
OS Synthetic.

```







Search completed: July 3, 2005, 02:38:07  
Job time : 202.647 secs

**This Page Blank (usps)**

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	130	100.0	452	4	US-09-263-692A-1		Sequence 1, Appli
2	89	68.5	89	4	US-09-263-692A-13		Sequence 13, Appli
3	60.4	46.5	1138	3	US-09-011-151-8		Sequence 8, Appli
4	60.4	46.5	1138	3	US-09-011-151-9		Sequence 9, Appli
5	59.6	45.8	157	3	US-09-545-244A-9		Sequence 9, Appli
6	56.2	43.2	5033	1	US-08-038-768A-1		Sequence 1, Appli
7	52	40.0	6395	2	US-08-687-559-2		Sequence 2, Appli
8	52	40.0	6395	3	US-09-259-741-1		Sequence 1, Appli
9	52	40.0	6395	3	US-09-037-751-1		Sequence 1, Appli
10	52	40.0	6395	3	US-09-466-422-1		Sequence 1, Appli
11	52	40.0	6395	4	US-09-401-415-2		Sequence 2, Appli
12	52	40.0	6395	4	US-09-962-527-1		Sequence 1, Appli
13	52	40.0	6425	3	US-09-259-741-3		Sequence 3, Appli
14	52	40.0	6425	3	US-09-037-751-3		Sequence 3, Appli
15	52	40.0	6425	3	US-09-466-422-3		Sequence 3, Appli
16	52	40.0	6425	4	US-09-962-527-3		Sequence 3, Appli
17	52	40.0	6439	3	US-09-259-741-2		Sequence 2, Appli
18	52	40.0	6439	3	US-09-037-751-2		Sequence 2, Appli
19	52	40.0	6439	3	US-09-466-428-2		Sequence 2, Appli
20	52	40.0	6439	4	US-09-962-527-2		Sequence 2, Appli
21	52	40.0	6446	3	US-09-259-741-5		Sequence 5, Appli
22	52	40.0	6446	3	US-09-037-751-5		Sequence 5, Appli
23	52	40.0	6446	3	US-09-466-422-5		Sequence 5, Appli
24	52	40.0	6446	3	US-09-962-527-5		Sequence 5, Appli
25	52	40.0	6475	3	US-09-259-741-4		Sequence 4, Appli
26	52	40.0	6475	3	US-09-037-751-4		Sequence 4, Appli
27	52	40.0	6475	3	US-09-466-422-4		Sequence 4, Appli



CITY: Des Moines  
STATE: Iowa  
COUNTRY: United States  
ZIP: 50309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS, Microsoft Windows  
SOFTWARE: Microsoft Windows No. 5569828epad  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/038,768A  
FILING DATE: 19930324  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Roth, Michael J.  
REGISTRATION NUMBER: 29,342  
REFERENCE/DOCKET NUMBER: 0235 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (515) 245-3594  
TELEFAX: (515) 245-3634  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5033 base pairs  
TYPE: nucleotide  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: synthetic DNA  
DESCRIPTION: transformation plasmid pPHI1406  
HYPOTHETICAL: No  
US-08-038-768A-1

Query Match 43.2%; Score 56.2; DB 1; Length 5033;  
Best Local Similarity 83.2%; Pred. No. 1.6e-05;  
Matches 89; Conservative 0; Mismatches 13; Indels 5; Gaps 2;

QY 12 AGGAAGTTCATTTCATTGGAATGAC---ACGGTTGTGCTATTCTCAACAATTACCAAC 68  
|||||  
DB 1165 AGGAGTTCATTTCATTGGAATGACGCTGAGCTGAGCTTATTTCACAAATTACCAAC 1224  
|||||  
QY 69 AACCAACAAACAAACAAACATTATACAAATTACTATTATTACAAATTACA 115  
|||||  
DB 1225 AACCAACAAACAAACAAACAT--TACAATTACTATTATTACAAATTACA 1269  
|||||

RESULT 7  
US-08-687-559-2  
Sequence 2, Application US/08687559  
Patent No. 5955647  
GENERAL INFORMATION:  
APPLICANT: Fitchen, John H.  
APPLICANT: Beachy, Roger N.  
TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO  
PRODUCE PEPTIDES AND PROTEINS  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,559  
FILING DATE: No. 5955647ember 18, 1996  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01467  
FILING DATE: 03-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Learn, June M.  
REGISTRATION NUMBER: 31,238  
REFERENCE/DOCKET NUMBER: 07302/011001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6395 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
IMMEDIATE SOURCE:  
CLONE: TWV  
US-08-687-559-2

Query Match 40.0%; Score 52; DB 2; Length 6395;  
Best Local Similarity 85.4%; Pred. No. 0.00019;  
Matches 70; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCAATTTCTCAACAATTACCAACAACAACAACAACAACAACAATTATACAAATTACTATT 105  
|||||  
DB 1 GTATTTTACAAATTACCAACAACAACAACAACAACAACAATTATACAAATTACTATT 58  
|||||  
QY 106 TACAATTACATCTAGATAACA 127  
|||||  
DB 59 TACAATTACAAATGGCATAACA 80  
|||||

RESULT 8  
US-09-259-741-1  
Sequence 1, Application US/09259741  
Patent No. 6033895  
GENERAL INFORMATION:  
APPLICANT: GARGER, STEPHEN  
APPLICANT: HOLTZ, R. BARRY  
APPLICANT: MCCULLOCH, MICHAEL  
APPLICANT: TURPEN, THOMAS  
TITLE OF INVENTION: A PROCESS FOR ISOLATING AND  
PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT  
TISSUES  
TITLE OF INVENTION: SOURCES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howrey & Simon  
STREET: 1299 Pennsylvania Avenue N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/259,741  
FILING DATE: February 25, 1999  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/037,751  
FILING DATE: March 10, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Halluin, Albert P  
REGISTRATION NUMBER: 25,277  
REFERENCE/DOCKET NUMBER: 00801.0140.US01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-463-8100  
TELEFAX: 650-463-8400





RESULT 11  
US-09-401-415-2  
; Sequence 2, Application US/09401415  
; Patent No. 6503732  
; GENERAL INFORMATION:  
; APPLICANT: THE SCRIPPS RESEARCH INSTITUTE  
; TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO  
; OVERPRODUCE PEPTIDES AND PROTEINS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Spensley Horn Jubas & Lubitz  
; STREET: 1880 Century Park East, Suite 500  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/401.415  
; FILING DATE: 21-Sep-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/01467  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bostich, June M.  
; REGISTRATION NUMBER: 31,238  
; REFERENCE/DOCKET NUMBER: FD-4074  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 455-5100  
; TELEFAX: (619) 455-5110  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; IMMEDIATE SOURCE:  
; CLONE: TMV  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..6395  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-401-415-2

Query Match 40.0%; Score 52; DB 4; Length 6395;  
Best Local Similarity 85.4%; Pred. No. 0.00019;  
Matches 70; Conservative 0; Mismatches 10; Indels 2; Gaps 1;  
QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTACTACTATT 105  
Db 1 GTATTTTCAACAATTACCAACAACAACAACAACAACAACAACAACAACAACAACAATTACTACTATT 58  
QY 106 TACAATTACATCTAGATAACA 127  
Db 59 TACAATTACATCTAGATAACA 80

RESULT 12  
US-09-962-527-1  
; Sequence 1, Application US/09962527  
; Patent No. 6740740  
; GENERAL INFORMATION:  
; APPLICANT: GARGER, STEPHEN  
; ADDRESSEE: HOLTZ, R. BARRY  
; CITY: McCULLOCH, MICHAEL

TURPEN, THOMAS  
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND  
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES  
; FROM PLANT SOURCES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howrey & Simon  
; STREET: 1299 Pennsylvania Avenue N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/962,527  
; FILING DATE: 24-Sep-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/037,751  
; FILING DATE: 10-march-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Halluin, Albert P  
; REGISTRATION NUMBER: 25,277  
; REFERENCE/DOCKET NUMBER: 00801.0140.999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-463-8109  
; TELEFAX: 650-463-8400  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: Genomic RNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-962-527-1

Query Match 40.0%; Score 52; DB 4; Length 6395;  
Best Local Similarity 64.6%; Pred. No. 0.00019;  
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;  
QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTACTACTATT 105  
Db 1 GUUUUUUACAACAUAUCCAAACAACAACAACAACAACAACAACAACAUAU--UACAUAUUAU 58  
QY 106 TACAATTACATCTAGATAACA 127  
Db 59 UACAUAUUAUUAUCCAAACAACAACAACAACAACAACAACAACAUAU 80

RESULT 13  
US-09-259-741-3  
; Sequence 3, Application US/09259741  
; Patent No. 6033895  
; GENERAL INFORMATION:  
; APPLICANT: GARGER, STEPHEN  
; ADDRESSEE: HOLTZ, R. BARRY  
; CITY: McCULLOCH, MICHAEL  
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND  
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT  
; SOURCES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howrey & Simon  
; STREET: 1299 Pennsylvania Avenue N.W.  
; CITY: Washington  
; STATE: DC

```
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/259,741
; APPLICATION NUMBER: US/09/259,741
; FILING DATE: February 25, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: March 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8100
; TELEFAX: 650-463-8400
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; US-09-259-741-3

Query Match 40.0%; Score 52; DB 3; Length 6425;
Best Local Similarity 64.6%; Pred. No. 0.0002;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAATTATACATTACTATT 105
Db 1 GUUUUUUACAACAUAUCCAAACAACAACAACAACAACAACAUA--UACAUAUUAUUU 58

QY 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAUAUACAUAUGGCAUACACA 80

RESULT 14
US-09-037-751-3
; Sequence 3, Application US/09037751
; Patent No. 6037456
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,751
; FILING DATE: 10-MAR-1998
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; US-09-037-751-3

Query Match 40.0%; Score 52; DB 3; Length 6425;
Best Local Similarity 64.6%; Pred. No. 0.0002;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAATTATACATTACTATT 105
Db 1 GUUUUUUACAACAUAUCCAAACAACAACAACAACAACAACAUA--UACAUAUUAUUU 58

QY 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAUAUACAUAUGGCAUACACA 80

RESULT 15
US-09-466-422-3
; Sequence 3, Application US/09466422
; Patent No. 6303779
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/466,422
; FILING DATE: 17-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
```

```
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6425 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-466-422-3

Query Match          40.0%; Score 52; DB 3; Length 6425;
Best Local Similarity 64.6%; Pred. No. 0.0002;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

Qy 46 GTCAATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1  GUUUUUUACAACAUAUACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 58

Qy 106 TACAATTACATCTAGATAAACA 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 UACAUAUACAUAUGGCAUACACA 80
```

Search completed: July 3, 2005, 04:55:26  
Job time : 59.5294 secs

**This Page Blank (usp...)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 03:19:18 ; Search time 266.765 Seconds  
(without alignments)  
3056.566 Million cell updates/sec

Title: US-10-814-858A-2  
Perfect score: 130  
Sequence: 1 tcaatatataggaggttc.....ttacatctagataaacaatg 130

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/1/pubpna/PCUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*

10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*

12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*

13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*

14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*

15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*

17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*

18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*

19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:\*

20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*

21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*

22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*

23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:\*

24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*

25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*

26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63.6	48.9	2186	US-10-784-295-13	Sequence 13, Appl
2	54.8	42.2	368	US-09-331-631A-29	Sequence 29, Appl
3	54.8	42.2	368	US-10-147-095-29	Sequence 29, Appl
4	52	40.0	6395	US-09-962-527-1	Sequence 2, Appl
5	52	40.0	6395	US-10-338-592-2	Sequence 2, Appl
6	52	40.0	6395	US-10-828-029-1	Sequence 1, Appl
7	52	40.0	6425	US-09-962-527-3	Sequence 3, Appl

8	52	40.0	6425	19	US-10-828-029-3	Sequence 3, Appl
9	52	40.0	6439	10	US-09-962-527-2	Sequence 2, Appl
10	52	40.0	6439	19	US-10-828-029-2	Sequence 2, Appl
11	52	40.0	6446	10	US-09-962-527-5	Sequence 5, Appl
12	52	40.0	6446	19	US-10-828-029-5	Sequence 5, Appl
13	52	40.0	6475	10	US-09-962-527-4	Sequence 4, Appl
14	52	40.0	6475	19	US-10-828-029-4	Sequence 4, Appl
15	52	40.0	7926	14	US-10-119-330-1	Sequence 1, Appl
16	52	40.0	7926	18	US-10-632-240-1	Sequence 1, Appl
17	52	40.0	7926	21	US-10-976-698-1	Sequence 1, Appl
18	52	40.0	10600	16	US-10-356-708-1	Sequence 1, Appl
19	52	40.0	10600	19	US-10-280-913A-1	Sequence 1, Appl
20	52	40.0	10600	19	US-10-684-134-1	Sequence 1, Appl
21	52	40.0	10600	19	US-10-637-758-1	Sequence 1, Appl
22	52	40.0	10624	16	US-10-356-708-2	Sequence 2, Appl
23	52	40.0	10624	19	US-10-280-913A-2	Sequence 2, Appl
24	52	40.0	10624	19	US-10-684-134-2	Sequence 2, Appl
25	52	40.0	10624	19	US-10-637-758-2	Sequence 2, Appl
26	52	40.0	11222	19	US-10-679-620-73	Sequence 73, Appl
27	52	40.0	11641	9	US-09-993-059-33	Sequence 33, Appl
28	52	40.0	11641	15	US-10-103-327-33	Sequence 33, Appl
29	52	40.0	11641	18	US-10-684-300-13	Sequence 13, Appl
30	52	40.0	11641	18	US-10-684-349-13	Sequence 13, Appl
31	52	40.0	11641	20	US-10-851-388-33	Sequence 33, Appl
32	52	40.0	11641	22	US-10-984-389-33	Sequence 33, Appl
33	51.8	39.8	1824	9	US-09-930-329-1	Sequence 1, Appl
34	51.8	39.8	1824	9	US-09-930-342-1	Sequence 1, Appl
35	51.8	39.8	7688	9	US-09-949-317-27	Sequence 27, Appl
36	51.8	39.8	7688	9	US-09-949-316-27	Sequence 27, Appl
37	51.8	39.8	7688	14	US-10-200-051-27	Sequence 27, Appl
38	50.4	38.8	7685	9	US-09-949-317-22	Sequence 22, Appl
39	50.4	38.8	7685	9	US-09-949-317-25	Sequence 25, Appl
40	50.4	38.8	7685	9	US-09-949-316-22	Sequence 22, Appl
41	50.4	38.8	7685	9	US-09-949-316-25	Sequence 25, Appl
42	50.4	38.8	7685	14	US-10-200-051-22	Sequence 22, Appl
43	50.4	38.8	7685	14	US-10-200-051-25	Sequence 25, Appl
44	50.4	38.8	7686	9	US-09-949-317-23	Sequence 23, Appl
45	50.4	38.8	7686	9	US-09-949-317-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1

US-10-784-295-13

; Sequence 13, Application US/10784295

; Publication No. US2005003221A1

; GENERAL INFORMATION:

; APPLICANT: Shaaltiel, Yoseph

; TITLE OF INVENTION: CELL/TISSUE CULTURING DEVICE, SYSTEM AND METHOD

; FILE REFERENCE: 27557

; CURRENT APPLICATION NUMBER: US/10/784,295

; CURRENT FILING DATE: 2004-02-24

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 13

; LENGTH: 2186

; TYPE: DNA

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Nucleic acid encoding recombinant GCD fused to signal peptides

; NAME/KEY: misc feature

; LOCATION: (2181)..(2181)

; OTHER INFORMATION: n is a, c, g, or t

US-10-784-295-13

Query Match 48.9%; Score 63.6; DB 21; Length 2186;

Best Local Similarity 87.7%; Pred. No. 7.4e-06;

Matches 93; Conservative 0; Mismatches 9; Indels 4; Gaps 2;

QY 12 AGCAAGTTCATTTCATTTGGAATGACA--CGTGTTCATTTCTCAACAATACCAACA 69

|||||

```
Db 312 AGGAAGTTCATTCTTTCGAGAGGACAGGCTTCTTGAGATCCTTCAACAATTACCAACA 371
QY 70 ACAACAAACAAACAAACAAATTTACAAATTTACTATTATTTACAAATTACA 115
Db 372 ACAACAAACAAACAAACAAAT--TACAAATTACTATTATTTACAAATTACA 415

RESULT 2
US-09-331-631A-29
; Sequence 29, Application US/09331631A
; Patent No. US20020168392A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie L.
; TITLE OF INVENTION: ANTIMICROBIAL PROTEINS
; FILE REFERENCE: CULN23.001APC
; CURRENT APPLICATION NUMBER: US/09/331,631A
; CURRENT FILING DATE: 1999-06-21
; PRIOR APPLICATION NUMBER: PCT/AU97/00874
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: AU PO 4275
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A synthetic nucleotide sequence which can be used
; OTHER INFORMATION: for the expression and secretion of MiAMP2c.
; OTHER INFORMATION: containing the leader sequence from SEQ ID NO:11
; OTHER INFORMATION: and SEQ ID NO:5.
; NAME/KEY: CDS
; LOCATION: (103)...(333)
US-09-331-631A-29
Query Match 42.2%; Score 54.8; DB 9; Length 368;
Best Local Similarity 76.9%; Pred. No. 0.00064;
Matches 80; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

QY 27 TTTGGAATGGACACGCTGTTCTCAATTTCTCAACAATTACCAACAAACAAACAAACA 86
Db 4 TCTAGAGCGCGCGTCGACTATTTTACAACAATTACCAACAAACAAACAAACAAACA 63

QY 87 ACATTATACAATTACTATTATTTACAATTACATCTAGATAAACAATG 130
Db 64 ACAT--TACAATTACTATTACAATTACAGGATCCACAACAATG 105

RESULT 3
US-09-331-631A-29
; Sequence 29, Application US/10147095
; Publication No. US20030171274A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie L.
; TITLE OF INVENTION: ANTIMICROBIAL PROTEINS
; FILE REFERENCE: CULN23.001APC
; CURRENT APPLICATION NUMBER: US/10/147,095
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US/09/331,631A
; PRIOR FILING DATE: 1999-06-21
; PRIOR APPLICATION NUMBER: PCT/AU97/00874
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: AU PO 4275
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A synthetic nucleotide sequence which can be used
; OTHER INFORMATION: for the expression and secretion of MiAMP2c.
; OTHER INFORMATION: containing the leader sequence from SEQ ID NO:11
; OTHER INFORMATION: and SEQ ID NO:5.
; NAME/KEY: CDS
; LOCATION: (103)...(333)
US-09-331-631A-29
Query Match 42.2%; Score 54.8; DB 9; Length 368;
Best Local Similarity 76.9%; Pred. No. 0.00064;
Matches 80; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

QY 27 TTTGGAATGGACACGCTGTTCTCAATTTCTCAACAATTACCAACAAACAAACAAACA 86
Db 4 TCTAGAGCGCGCGTCGACTATTTTACAACAATTACCAACAAACAAACAAACAAACA 63

QY 87 ACATTATACAATTACTATTATTTACAATTACATCTAGATAAACAATG 130
Db 64 ACAT--TACAATTACTATTACAATTACAGGATCCACAACAATG 105

RESULT 4
US-09-962-527-1
; Sequence 1, Application US/09962527
; Publication No. US20030049813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: 10-march-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
```

```
; SEQ ID NO 29
; LENGTH: 368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A synthetic nucleotide sequence which can be used
; OTHER INFORMATION: for the expression and secretion of MiAMP2c.
; OTHER INFORMATION: containing the leader sequence from SEQ ID NO:11
; OTHER INFORMATION: and SEQ ID NO:5.
; NAME/KEY: CDS
; LOCATION: (103)...(333)
US-10-147-095-29
Query Match 42.2%; Score 54.8; DB 16; Length 368;
Best Local Similarity 76.9%; Pred. No. 0.00064;
Matches 80; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

QY 27 TTTGGAATGGACACGCTGTTCTCAATTTCTCAACAATTACCAACAAACAAACAAACA 86
Db 4 TCTAGAGCGCGCGTCGACTATTTTACAACAATTACCAACAAACAAACAAACAAACA 63

QY 87 ACATTATACAATTACTATTATTTACAATTACATCTAGATAAACAATG 130
Db 64 ACAT--TACAATTACTATTACAATTACAGGATCCACAACAATG 105

RESULT 4
US-09-962-527-1
; Sequence 1, Application US/09962527
; Publication No. US20030049813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: 10-march-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
```

```
;
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-962-527-1

Query Match          40.0%; Score 52; DB 10; Length 6395;
Best Local Similarity 64.6%; Pred. No. 0.007;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GUUUUUUACAACAUAUACCAACAACAACAACAACAACAACAACAACAUA--UACAUAUUAUUU 58

QY 106 TACAATTACATCTAGATAAACA 127
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 59 UACAUAUUAUACAUGGCAUACACA 80

RESULT 5
US-10-338-592-2
; Sequence 2, Application US/10338592
; Publication No. US20030208792A1
; GENERAL INFORMATION:
; APPLICANT: Fitch, John H.
; TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO
; FILE REFERENCE: TSRI 312.1C1
; CURRENT APPLICATION NUMBER: US/10/338,592
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 09/401,415
; PRIOR FILING DATE: 1999-09-21
; PRIOR APPLICATION NUMBER: US 08/687,559
; PRIOR FILING DATE: 1996-11-18
; PRIOR APPLICATION NUMBER: PCT/US95/01467
; PRIOR FILING DATE: 1995-02-03
; PRIOR APPLICATION NUMBER: US 08/192,477
; PRIOR FILING DATE: 1994-02-03
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6395
; TYPE: DNA
; ORGANISM: Tobacco mosaic virus
US-10-338-592-2

Query Match          40.0%; Score 52; DB 17; Length 6395;
Best Local Similarity 85.4%; Pred. No. 0.007;
Matches 70; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GTATTTTACAACAATTACCAACAACAACAACAACAACAACAACAACAAT--TACAATTACTATT 58

QY 106 TACAATTACATCTAGATAAACA 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 TACAATTACAATGGCATAACA 80

RESULT 6
US-10-828-029-1
; Sequence 1, Application US/10828029
; Publication No. US20040171813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCCULLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```
;
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/828,029
; FILING DATE: 20-Apr-2004
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; APPLICATION NUMBER: 09/037,751
; FILING DATE: 10-march-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-828-029-1

Query Match          40.0%; Score 52; DB 19; Length 6395;
Best Local Similarity 64.6%; Pred. No. 0.007;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GUUUUUUACAACAUAUACCAACAACAACAACAACAACAACAACAACAUA--UACAUAUUAUUU 58

QY 106 TACAATTACATCTAGATAAACA 127
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 59 UACAUAUUAUACAUGGCAUACACA 80

RESULT 7
US-09-962-527-3
; Sequence 3, Application US/09962527
; Publication No. US20030049813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCCULLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```
;
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: 10-march-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-962-527-3
Query Match 40.0%; Score 52; DB 10; Length 6425;
Best Local Similarity 64.6%; Pred. No. 0.007;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
Db 1 GUUUUUUACAACAAUUUACCAACAACAACAACAACAACAACAACA--UACAUAUUAUAUU 58

QY 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAUAUUAACAAUGGCAUACACA 80

RESULT 8
US-10-828-029-3
; Sequence 3, Application US/10828029
; Publication No. US20040171813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCCULLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/828,029
; FILING DATE: 20-Apr-2004
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; APPLICATION NUMBER: 09/037,751

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
Db 1 GUUUUUUACAACAAUUUACCAACAACAACAACAACAACAACAACA--UACAUAUUAUAUU 58

QY 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAUAUUAACAAUGGCAUACACA 80

RESULT 9
US-09-962-527-2
; Sequence 2, Application US/09962527
; Publication No. US20030049813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCCULLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: 10-march-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-828-029-3
Query Match 40.0%; Score 52; DB 19; Length 6425;
Best Local Similarity 64.6%; Pred. No. 0.007;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
Db 1 GUUUUUUACAACAAUUUACCAACAACAACAACAACAACAACAACA--UACAUAUUAUAUU 58

QY 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAUAUUAACAAUGGCAUACACA 80

RESULT 9
US-09-962-527-2
; Sequence 2, Application US/09962527
; Publication No. US20030049813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCCULLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: 10-march-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-828-029-3
Query Match 40.0%; Score 52; DB 19; Length 6425;
Best Local Similarity 64.6%; Pred. No. 0.007;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
Db 1 GUUUUUUACAACAAUUUACCAACAACAACAACAACAACAACAACA--UACAUAUUAUAUU 58

QY 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAUAUUAACAAUGGCAUACACA 80
```









This page blank (USP12)